

遠隔コミュニケーションにおける 発話衝突低減手法

平成24年度

玉木 秀和

目次

第1章	序論	1
1.1	研究の背景と目的	2
1.2	研究の概要	2
1.3	本研究の位置づけ	4
1.4	本研究の学術的貢献	5
1.5	本論文の構成	6
第2章	遠隔コミュニケーションにおける話者交替	8
2.1	コミュニケーションとは	9
2.2	話者交替	10
2.2.1	対面コミュニケーションにおける話者交替	11
2.2.2	話者交替のルール	11
2.2.3	遠隔コミュニケーションにおける話者交替	12
2.3	非言語情報	13
2.3.1	非言語情報の役割	13
2.3.2	非言語情報の分類	14
2.4	対面コミュニケーションの支援に関する先行研究	16
2.4.1	適切な話題の提供支援	16
2.4.2	人とエージェントとの会話	17
2.5	遠隔コミュニケーションの支援に関する先行研究	17
2.5.1	テレプレゼンスを目指した遠隔コミュニケーションシステム	18
2.5.2	アバタを用いるシステム	21
2.5.3	その他のとりくみ	22
第3章	音声遅延と発話衝突確率および精神的ストレスの関係	24
3.1	遠隔コミュニケーションにおける話者交替の弊害	25
3.2	音声遅延による発話衝突のメカニズム	25
3.3	実験手順	26
3.4	実験結果	29
3.4.1	発話衝突確率	29
3.4.2	精神的ストレス	34
3.4.3	会話の効率	34

3.4.4	考察	35
3.5	快適な Web 会議システム構築へ向けた課題	38
第 4 章	発話衝突低減手法を確立するためのアプローチ	41
4.1	分析	42
4.1.1	Web 会議における発話衝突の例	42
4.1.2	Web 会議における予備動作の使われ方の分析	45
4.2	発話衝突を低減するための基本コンセプト	47
第 5 章	次話者候補提示法	49
5.1	次話者候補提示手法の提案	50
5.2	オズの魔法使い実験による検証	50
5.3	プロトタイプ実装	51
5.3.1	予備動作の検知	51
5.3.2	次話者候補の選定方法	52
5.4	実験	55
5.4.1	目的	55
5.4.2	手順	55
5.4.3	結果	56
5.5	考察	58
5.6	課題	61
第 6 章	発話欲求伝達手法	62
6.1	発話欲求伝達手法の提案	63
6.2	発話欲求伝達モジュールの実装	64
6.2.1	予備動作候補検知部	65
6.2.2	発話欲求推定部	67
6.2.3	発話欲求伝達部	68
6.3	評価実験	69
6.3.1	概要	69
6.3.2	手順	69
6.3.3	結果	70
6.3.4	考察	71
第 7 章	結論	74
	謝辞	77
	参考文献	79
	論文目録	85

目 次

1.1	本研究の位置づけ	5
3.1	話者交替時の認知・行動モデル	26
3.2	音声遅延による発話衝突への影響	27
3.3	音声遅延会議システム	28
3.4	7分間での平均発話回数の比較	30
3.5	発話衝突確率	30
3.6	質問紙評価項目「音声遅延をストレスに感じる」の結果	31
3.7	質問紙評価項目「誰が話しているか分からないことをストレスに感じる」の結果	31
3.8	質問紙評価項目「誰が話し始めそうか分からないことをストレスに感じる」の結果	31
3.9	質問紙評価項目「いつ話し始めていいか分からないことをストレスに感じる」の結果	32
3.10	質問紙評価項目「話し始めが他の人とぶつかることをストレスに感じる」の結果	32
3.11	質問紙評価項目「他の参加者が自分の話を聞いているか分からない」の結果	32
3.12	質問紙評価項目「沈黙が多い」の結果	33
3.13	質問紙評価項目「話に割り込めないことをストレスに感じる」の結果	33
3.14	質問紙評価項目「盛り上がらない」の結果	33
3.15	質問紙評価項目「相手の存在感が薄い」の結果	34
3.16	発話開始間隔の比較	38
3.17	発話音声再生開始間隔の比較	39
4.1	予備動作の分類	43
4.2	複数の参加者が同時に発話して衝突する様子	44
4.3	発話の切れ目に割り込もうとして衝突する様子	44
4.4	予備動作後の発話の非衝突確率	46
4.5	Web 会議における予備動作回数と発話の関係	47
4.6	話者交替時の認知・行動モデルにおいて本提案コンセプトが支援するステップ	47
5.1	オズの魔法使い実験の結果	51
5.2	予備動作毎のスコアの推移と発話可能性ポイント	53

5.3	プロトタイプシステム実行画面	54
5.4	発話可能性ポイントの推移と次話者候補選択の例	54
5.5	動作非教示群における，次話者候補提示有無による発話回数の比較	56
5.6	動作教示群における，次話者候補提示有無による発話回数の比較	56
5.7	次話者候補提示有無による発話回数増加量の比較	57
5.8	動作非教示群における，次話者候補提示枠有無による発話衝突確率の比較	58
5.9	動作教示群における，次話者候補提示枠有無による発話衝突確率の比較	59
5.10	時話者候補提示有無による発話衝突確率減少度の比較	59
5.11	発話欲求が生じてから発話するまでのステップ	60
5.12	次話者候補提示枠表示後発話確率と発話衝突確率減少度の関係	60
6.1	発話欲求を伝達可能な遠隔会議システム実行画面	64
6.2	会議システム全体の構成図	65
6.3	キャプチャされた参加者の関節の3次元位置	66
6.4	予備動作の検知方法	67
6.5	参加者映像	68
6.6	ターン数の比較	71
6.7	発話衝突確率の比較	71
6.8	「発話が衝突したと感じた度合い」の比較	72
6.9	話者交替において提案手法を適用した際の認知・行動モデル	73

表 目 次

2.1	非言語情報の分類 [14, 47]	14
3.1	Friedman の検定結果	35
3.2	多重比較結果「音声遅延をストレスに感じる」	36
3.3	多重比較結果「話し始めが他の人とぶつかることをストレスに感じる」	36
3.4	アイデア創出数	36
4.1	動作後の発話確率	46
5.1	発話の予備動作のスコア付け	53
6.1	被験者グループごとの実施した条件順序	70

第1章 序論

1.1 研究の背景と目的

環境負荷低減や出張費削減などが求められる中、遠隔会議システムの需要が増加してきている。遠隔会議システムの中でも Web 会議システムは、以下の理由からユーザに人気があり、市場も成長傾向にある [54]。1 つ目の理由としては複数の離れた拠点にいる会議の参加者が自席のデスクトップパソコンやノートパソコンから容易に参加できる利用のしやすさが挙げられる。2 つ目の理由としては、インターネットに接続したパソコンに安価な Web カメラとヘッドセットを接続し、ソフトウェアをインストールすればすぐに利用できる導入のしやすさが挙げられる。

しかし、複数の拠点を結んで行う Web 会議において、参加者が互いの様子を正確に読み取ることは困難である [65]。その原因として以下の点が考えられる。個々の参加者の映像はディスプレイを分割した領域に表示されるため、表示サイズが多くの場合小さくなってしまうこと、帯域の保証されていないネットワーク上で Web 会議を実施する場合、映像や音声は遅延して伝達されること、また、ある 1 つの拠点の映像、音声、通信などの質が低い場合、それが他の拠点の環境に影響を及ぼしてしまうことである。

したがって、Web 会議において話者交替が頻繁に起こる会議を実施した場合、2 人以上の参加者が意図せず同時に発話を開始する発話衝突が多く生じる [73]。会議はその性質から、伝達会議、調整会議、決定会議、創造会議の 4 つに分類される [59] が、創造会議は新しいアイデアを考え出す会議を指し、その知的生産性を高めるためには会議参加者各々がアイデアや意見を活発に出すことが重要だとされている。そのため、創造会議では話者交替が頻繁に起こり [18]、発話衝突が多く起こる Web 会議で実施することは難しいと考えられる。

そこで本研究は、3 人以上の参加者がいる Web 会議での発話衝突を低減することにより、自席のデスクトップパソコンやノートパソコンからでも、創造会議を快適に行える遠隔会議環境を実現することを目的とする。参加者が 2 人である場合、話者交替において話者の入れ替わる順序は単純である。しかし参加者が 3 人以上いる場合、ある 1 人の発話が終了した時点で、次に発話を開始する可能性のある参加者は 2 人以上いることとなる。そのため発話衝突確率は増え、円滑な話者交替を行うことはさらに困難になると考えられる。

本論文では、1 つの拠点に 1 人の参加者がいる場合に状況を絞って研究を行う。実際に Web 会議を行う際には、1 つの拠点に複数の参加者がいることがあるが、その場合、同じ拠点にいる参加者と、他の拠点にいる参加者とで、話者交替を行うときの状況が変わり、問題が複雑になると考えられる。そのため、まずは 1 つの拠点に 1 人の参加者がいる状況に絞って発話衝突の問題を分析し、研究を進める。

1.2 研究の概要

Web 会議システムを用いて創造会議を実施する場合、発話衝突が問題となる。話者交替が失敗する要素には発話衝突と沈黙があると言われており [11]、Sacks は発話が衝突す

ると発話を諦めて中断する傾向が高いことを指摘している [50]。発話衝突が頻繁に起こる環境では、素早く話者交替をする会議は行いづらく [73]、会議の生産性が下がる可能性がある。

本研究では、発話衝突を低減させる手法を模索するために、まずは発話衝突の原因とその影響について分析した。Web 会議において、参加者は映像よりも音声の情報に頼って会話をするといわれている [65]。そのため Web 会議における話者交替は、音声の品質の善し悪しによる影響を大きく受けると考えられる。音声の品質の中でも特に、発話衝突に大きな影響を与える要因として音声遅延に着目した。話者交替時の認知・行動モデルを定義し、それがどのように音声遅延の影響を受けて発話衝突が生じるかという仮説をたて、実験を行った。音声遅延量を変化させられる音声会議環境を用意し、音声遅延量と発話衝突、またそれらに応じて参加者の受ける精神的ストレスの関係を調べた。

本実験環境では、音声遅延量が 400msec を超える場合には発話衝突確率と精神的ストレスが高まる結果となった。音声遅延量が 0msec の条件と 200msec の条件では、発話衝突確率と精神的ストレスの指標に差は見られなかった。本研究が対象とする Web 会議の環境では音声遅延量が 400msec を超えることが多いため [61]、発話衝突確率を低減させることが望まれる。

そこで本論文では、次話者候補提示手法と発話欲求伝達手法から構成される発話衝突低減手法を提案する。これらの手法のコンセプトは、人が発話する前に表出する特徴的な非言語情報をシステムが検知し、他の参加者へ強調して伝達することである。

次話者候補提示手法は、システムが会議参加者各々の予備動作を検知し、その種類と頻度から、次に最も発話しそうな参加者を次話者候補として選定し、すべての参加者へ伝達する手法である。予備動作とは、人が発話をする前に行う特徴的な非言語情報のことである。人は対面したコミュニケーションにおいて、頷きや身体の動きなど、予備動作を活用して円滑な話者交替を行っている [32, 67] が、Web 会議においてはこれらの伝達が困難であると考えられる。システムは最も多く予備動作を行った参加者を、そのときの次話者候補として選定する。そして、次話者候補をすべての参加者に分かりやすくするために、選定された参加者の映像が強調して表示される。この手法を実現するプロトタイプシステムを実装して評価実験を行った結果、本提案手法により Web 会議において発話衝突確率を低減できる可能性が示された。

発話欲求伝達手法は、システムによって推定された発話欲求の度合いを随時参加者に提示することにより、参加者自身に予備動作を調整させ、発話欲求の推定精度を高める手法である。システムが参加者の発話欲求を正確に推定することはできないため、参加者がそれを補助することで推定精度を高められる手法を提案した。この手法を実現するプロトタイプシステムを実装して評価実験を行い、本提案手法を用いることで通常の Web 会議と比較して発話衝突確率を半分に以下に低減できることを確認した。

以上の結果から、本論文で提案した発話衝突低減手法により、Web 会議において発話衝突確率が低減できることが分かった。したがって、複数の離れた拠点にいる参加者が自席のデスクトップパソコンやノートパソコンから参加し、創造会議を快適に行える遠隔会議環境を実現できる可能性を示すことができた。

1.3 本研究の位置づけ

本研究は、Web 会議での発話衝突を低減することにより、自席のデスクトップパソコンやノートパソコンからでも、創造会議を快適に行える遠隔会議環境を実現することを目指している。本節では本研究の位置づけを整理する。図 1.1 は、既存の遠隔コミュニケーション支援システムの分類と、本研究の位置づけを示している。図の横軸はシステムの導入と利用の手軽さを表している。第一，四象限に位置するシステムは、導入や利用のために必要な機材や資金が少なく、利用する場所の自由度も高い。第二，三象限に位置するシステムは、専用の大型機材や専用のネットワーク回線を要するもの、専用に作り込まれた部屋でなくては利用できないものなどが含まれる。図の縦軸は会話のしやすさを表している。第一，二象限に位置するシステムは、参加者は互いの様子を鮮明に見ることができ、音声は聞こえやすく、伝送遅延が小さい。第三，四象限に位置するシステムは、映像から相手の様子を鮮明に見ることは困難で、音声は聞こえづらく、伝送遅延が大きい。

この図の第二象限に位置するシステムは画像や音声の質が高く会話がしやすいが、システム専用の部屋やディスプレイ、ネットワークを用意する必要があり、導入費用も高いものが多い。この領域にはテレプレゼンスを目指すシステムが位置する。テレプレゼンスとは、遠隔地にいる参加者があたかも目の前にいるかのような臨場感を与える技術のことである。この領域を対象としたシステムに関しては盛んに研究開発が行われている [6, 27]。

同図第四象限に位置するシステムは、導入や利用のために必要な機材が少なく、利用する場所の自由度も高いが、多くの場合画像や音声の品質が低く円滑な会話をするのが難しい。この領域には Web 会議システムが位置する。自席のデスクトップパソコンやノートパソコンに安価な Web カメラやヘッドセットを接続し、ソフトウェアをインストールすることですぐに導入して利用することができるが、遠隔地にいる相手の参加者の様子を読み取ることが難しい場合が多く、発話衝突が起こりやすい。

現在商用化されている Web 会議システムの中には発話権の遷移をサポートする機能として、マイクのオン・オフを切り替えることにより発話権を獲得、放棄するためのボタンや、挙手アイコンを表示させるためのボタンなどを実装しているものもある [39]。

一般的に、専用の大型機材や専用のネットワーク回線を要するもの、専用に作り込まれた部屋でなくては利用できないシステムでは会話はしやすくなるが導入と利用のしやすさが下がり、逆に Web 会議システムのように導入や利用のために必要な機材や資金が少なく、利用する場所の自由度が高いシステムは会話のしやすさは下がるというように、これまでの遠隔コミュニケーション支援システムは図 1.1 の網かけ上の領域にマッピングされることが多かった。

本研究が対象とするのは第二象限の、導入や利用のために必要な機材が少なく、利用する場所の自由度が高く、さらに会話のしやすい環境の実現を目指す研究領域である。同じくこの領域を対象とする先行研究としては、Web 会議においてアバタが替わりに非言語情報を表現するシステムが [21, 24, 31] 多く存在するが、実際のユーザの非言語情報とは無関係にアバタの非言語情報を生成しているものがほとんどである。そのため、ユーザの意図が正しく伝わらず、会話がしづらくなる可能性がある。

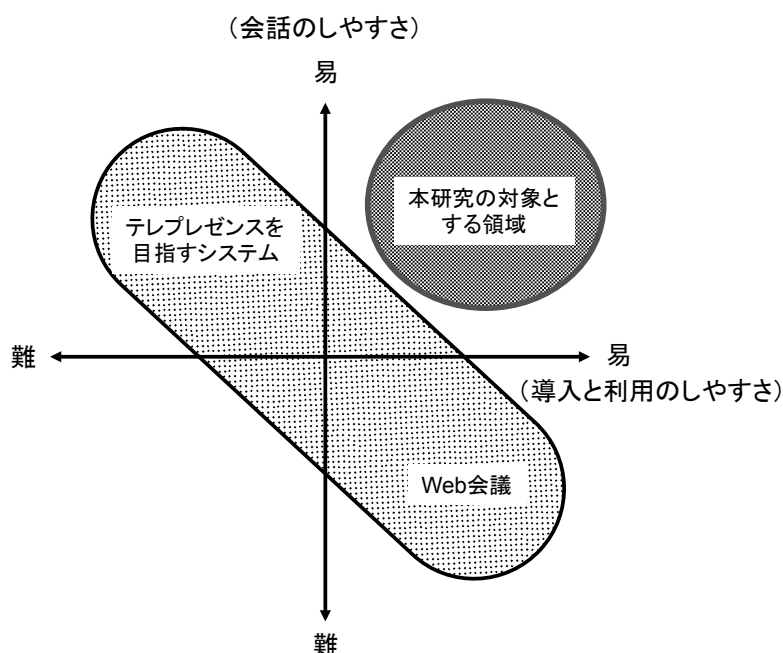


図 1.1: 本研究の位置づけ

そこで本研究では，Web 会議において参加者から表出する非言語情報を基に，話者交替に必要な情報を抽出して伝達することで円滑な話者交替を支援し，会話のしやすい環境を実現することを目指す．

1.4 本研究の学術的貢献

本研究の学術的な貢献は，遠隔コミュニケーションにおける話者交替時の認知・行動モデルを定義することと，遠隔コミュニケーションにおいて話者交替時の発話衝突を低減するための手法を提案することである．これまで対面コミュニケーションにおいては話者交替のルールが定義され，それに基づいて多くの議論がなされていた．しかし，遠隔コミュニケーションにおいてそれらのルールがどのように変化するか整理し，それを基に遠隔コミュニケーションならではの話者交替の問題を議論した例は少ない．本研究では，発話欲求，予備動作，観察，判断，行動の各ステップからなる，遠隔コミュニケーションにおける話者交替時の認知・行動モデルを定義する．今後発話衝突に限らず，話者交替の他の問題に関してもこのモデルに基づいて議論や研究を進めていくことができる．すなわち，発話衝突から復帰する，発話権の譲渡を円滑にする，意図せぬ沈黙を防ぐなどの課題の解決方法を考える際にも，このモデルの中で定義されたどのステップに問題が生じているかを捉え，それをどのように解決するかを筋道立てて考えることができる．

さらに，Web 会議での発話衝突を低減し，自席のデスクトップパソコンやノートパソ

コンからでも、新しいアイデアを考え出すことを目的とする創造会議を、快適に行える遠隔会議環境を実現するための手法を提案することで、導入や利用のために必要な機材が少なく、利用する場所の自由度が高く、さらに会話のしやすい環境を実現するための大きな一歩を示すことに貢献した。

1.5 本論文の構成

本論文は以下の7章から構成される。

第1章では、本研究の背景と目的を述べた後に研究の位置づけを整理し、本研究が学術的に貢献する点を述べた。

第2章では、本研究の対象である遠隔コミュニケーションについて整理する。まず始めにコミュニケーションの定義について整理する。次に本研究が対象とする話者交替に関して、先行研究を挙げて整理する。そしてコミュニケーションにおいて重要だとされている非言語情報について紹介し、具体的な対面コミュニケーション支援システム、遠隔コミュニケーション支援システムの先行研究を紹介する。

第3章では遠隔コミュニケーションにおいて、発話衝突の原因とその影響について調査する。まず話者交替時の認知・行動モデルを定義し、それが発話衝突の大きな原因であると考えられる音声遅延のある環境でどのような影響を受けるか仮説を立てる。次に音声会議環境で被験者実験を行い、音声遅延量を増減することで発話衝突確率がどのように変化するかを調べる。さらに音声遅延量と発話衝突確率と、会議参加者の受ける精神的ストレスの関係も明らかにする。得られた実験結果から、発話衝突と精神的ストレスの観点から会話に支障のない音声遅延量を明らかにする。またその基準を満たせない場合について、いくつかの解決策案を提示する。

第4章では、遠隔会議において発話衝突を低減するためのアプローチについて述べる。人は対面コミュニケーションにおいて予備動作を活用することで話者交替を円滑に行っている。そこでまず遠隔コミュニケーションを観察し、予備動作の活用度合いを調査する。そこから得られた、Web会議において予備動作が相手に認知されていないという気づきを基に、話者交替時の認知・行動モデルにおいて予備動作の伝達を支援することで発話衝突を低減するコンセプトを提案する。

第5章では、第4章で述べたコンセプトに基づき、遠隔会議において発話衝突を低減する手法を実装したプロトタイプシステムを構築し、その有効性を検証する。具体的には、会議参加者の予備動作をシステムが検知し、最も次に発話開始しそうな参加者を1人「次話者候補」として選出し、全参加者へ提示する。そしてプロトタイプを用いた被験者実験の結果から得られた、本手法の効果と課題を述べる。

第6章では、第5章の検証実験結果と考察から得られた知見を基に、発話衝突低減手法をさらに有効なものとするための改善をする。参加者の予備動作から必ずしも正確に発話欲求を予測できない問題を解決するために、システムに予測されている発話欲求を参加者へとフィードバックし、参加者各々が自身の予備動作を調整する仕組みを取り入れる。こ

の仕組みを実現するプロトタイプシステムを構築し，被験者実験によりその有効性を検証する．

第 7 章にて，本論文の結論を述べる．

第2章 遠隔コミュニケーションにおける 話者交替

本章では研究対象である遠隔コミュニケーションにおける話者交替について述べる．まずコミュニケーションの概念や定義をまとめ，本研究で特にフォーカスする話者交替について，対面コミュニケーションと遠隔コミュニケーションに分けて先行研究で述べられていることを整理する．次にコミュニケーションや話者交替を円滑に行うために重要な要素である非言語情報について述べる．そして対面／遠隔でのコミュニケーション支援システムについて先行研究を整理する．

2.1 コミュニケーションとは

「人間社会の基礎はコミュニケーションにある」[14]と言われるように，人間社会にとってコミュニケーションは必要不可欠なものである．このコミュニケーションは先行研究において，さまざまな視点から定義されている．

物的伝達機構である交通の問題に取り組んでいた Cooley は，20 世紀の初めにコミュニケーションを次のように定義した [7]．

コミュニケーションとは，それによって人間関係が成立し，発達するメカニズムを意味する．それは精神のすべてのシンボルであるとともに，空間を隔ててシンボルを運搬し，あるいは時間を経過した後までこれを保存する手段でもある．それは顔の表情，態度と身振り，声の調子，言葉，書字，印刷，鉄道，電信，電話，その他の時間と空間を克服するすべての事績を含む．

ただし，これは包括的な定義であり，曖昧さを含む．

Schramm はコミュニケーションの動機や意図に着目した [52] ．

コミュニケーションという言葉は，ラテン語の “communis”（共通，共有）から来ている．我々の間でのコミュニケーションとは，我々の間に共通性を成立させる．つまり，情報，思想，あるいは態度を共有しようとする試みである．

一方，深田は人がコミュニケーションを行う目的に着目し，下記のようにコミュニケーションを分類している [12] ．

(1) 課題解決

解決が必要な課題に関して相手から情報や援助を求める，相手との間で解決すべき課題に関して交渉や取り引きを行う等のことを目的としたコミュニケーション．

(2) 情報・知識の伝達

自分が知っている情報や知識を相手に伝達することを目的としたコミュニケーション．

(3) 情報・知識の獲得

自分自身に関する情報，相手に関する情報，他者・状況・環境等の外界に関する情報の獲得を目的としたコミュニケーション．

(4) 相手に対する影響力行使

相手の態度や行動を変えようと説得する，命令や強制等で相手を支配する，相手を騙す，相手を援助する等のことを目的としたコミュニケーション．

(5) 対人関係の形成・発展・維持

相手との間に友好的な対人関係を形成する，発展させる，維持する等のことを目的としたコミュニケーション．

(6) 娯楽の享受

退屈や孤独の気持ちを晴らすことを目的としたコミュニケーション．

なお，上記の目的はそれぞれが完全に独立しているのではなく，例えば「相手との対人関係を形成するという目的のために，相手に関する情報の獲得を目的とするコミュニケーションを行う」といった包摂関係にある項目も存在すると深田は述べている．

Dance と Larson は著書「人間のコミュニケーションの機能」の中で，過去に出版された研究書や論文等の文献から 126 にも及ぶコミュニケーションの定義をリストアップしている [9] ．

また，コミュニケーション学者 Wood は，「コミュニケーション」という言葉を定義する難しさには次の 3 つの理由があると主張している [71] ．

- (1) 我々はコミュニケーションについて，日常よく考えない傾向がある．コミュニケーションを当然のものとして受け入れてしまいがちなために，この実態，機能を良く分析せずに済ませてしまうからである．
- (2) コミュニケーション活動の見られる範囲が広すぎて，マス・コミュニケーションのメディア広告から，親密な対人関係のやり取りまですべての領域がカバーされるために，コミュニケーションを一つだけの定義で代表させることは難しい．
- (3) “コミュニケーション” という言葉が現代のキーワードとして広く用いられているため，コミュニケーションの中にコンピュータから人工衛星までをも含む傾向が，現代では顕著になっている．

このように様々なコミュニケーションの定義があるが，本論文で扱うコミュニケーションは，「2人以上の人が相互に意思や情報を伝えあうこと」とする．

2.2 話者交替

特に音声を用いてコミュニケーションを行う場合に，意思や情報を伝える側と受け取る側が交替することを話者交替という．本節では話者交替に関する既存研究の結果から得られている知見を整理する．

2.2.1 対面コミュニケーションにおける話者交替

対面したコミュニケーションにおいては、話者交替に関する研究が進んでいる [11, 28, 50] .
まず、基本的な用語を整理する [11] .

- ターン (turn)
発話権を得て話された1人の話者の一続きの発話 .
- ターン構成単位 (turn constructional unit; TCU)
ターンを構成しうる最小の単位であり、文、節、句、単語などである。「相槌」は除く。ターンは1つ以上のTCUから構成される .
- 話者交替適格場 (transition-relevance place; TRP)
各TCUの末尾に存在する話者交替にふさわしい場面。発話内容を細かく区切ったときの各々の「場面」を「場 (place)」と呼ぶ .
- 無標な重複 (unmarked overlap)
TRP付近に生じる音声的重複。この重複は聞き手が完結可能点を予測したうえで話出した結果生じるものとされ、話者交替に悪影響はない .
- 妨害 (interruption)
TRP以前のTCUに対してなされる重複。相手の発話権への侵害となる。Sacksの定義した話者交替のルール [50] (後述) に違反している .
- 発話衝突 (simultaneous talk / speech contention)
TRPにおいて、2人以上の参加者によって複数のTCUが同時に開始されることを指す .
- 話者交替の失敗
TRPにおいて、現話者が発話を継続する、もしくは他の参加者が発話を開始する場合を話者交替の成功とし、複数の参加者が発話を開始する場合に起こる衝突と、誰も発話せずに沈黙することを話者交替の失敗とする [28] .

2.2.2 話者交替のルール

Sacks は話者交替のルールを次のように定めた [50] .

- (1) 最初のTCUにおける最初のTRPにおいて、
 - (a) 「現話者による次話者選択 (言葉で明示するか、視線を移すといった行動)」が行われているなら、選択されたものが次のターンを取る権利と義務を得る .
 - (b) 「現話者による次話者選択」が行われていないなら、次の話者についての自己選択が可能になる。最初に話始めた者がターンを取る権利を得る .

(c) 「現話者による次話者選択」が行われておらず、他の者が自己選択も行わないなら、現話者が話し続けることが可能であるが、義務的ではない。

- (2) 最初の TRP においてルール (1)-(a) や (1)-(b) が作動せず、(1)-(c) に従って現話者が話し続けているならば、以降の TRP において、(1)-(a) ~ (1)-(c) が話者交替が生じるまで繰り返される。

上述のルール (1)-(b) において、2 人以上の参加者が同時に発話を開始する場合、発話衝突が生じるが、対面コミュニケーションにおいてそれは頻繁には起こらないと Sacks は述べている [50]。

2.2.3 遠隔コミュニケーションにおける話者交替

遠隔コミュニケーション時の参加者の挙動について研究した例はいくつも存在している [17, 51]。国民による遠隔コミュニケーションの特徴の違いについて調査したものや [53]、男女カップルの通話中の行動を調査したもの [46]、ネゴシエーションなどの特定のシーンでのコミュニケーションの仕方について調査したものなど [10, 64]、多岐にわたる。テキストチャットを用いたコラボレーションについて調査した先行研究では、高いパフォーマンスを上げるチーム内では発言が多く、議論は深く、ブレインストーミングにかかる時間は短いと報告されている [41]。ここでは特に、本研究のテーマである遠隔コミュニケーションでの話者交替について既に研究されている例を紹介する。

Web 会議システムでは、参加者映像が升目状に配置されたデザインが主流である。Sellen はこの升目状のデザインの会議システムと、対面会議、Hydra[3] を使用した場合の会話の構造を比較する実験を行った [55]。そして、これらの条件間で、同時発話や妨害の起こる確率などに差異が見られたことを報告した。また同時に、使用する会議システムにより会話の偏りに差異が見られないことも報告した。徳らは 1 つのディスプレイ上で升目状に表示される参加者映像は小さく、各々の様子をよく見ることができないと述べた [65]。著者らのこれまでの調査 [61] では、Web 会議における会話では、発話衝突が頻繁に起こることが分かっている。Web 会議システムを使用して 4 地点から接続して創造会議を実施したところ、すべての発話のうち約 20 % の発話が衝突した。

音声情報が損なわれる要因には遅延、サンプリングレート、通信帯域、エコーなどが挙げられるが、遠隔で会議を行う場合に、最も大きく、なくすことができない要因は音声遅延である。鎧沢らは、2 地点間の映像コミュニケーションを行う際に、交互に数字を読み上げるような極端に速い話者交替を必要とする会話は、音声遅延量が増加するほど、会話が妨害される度合いも増すが、特別速い話者交替を必要としない会話であれば往復 500msec まで音声遅延が生じて問題はないと報告した [73]。ITU-T 勧告 G.114 は遠隔コミュニケーションにおける伝送遅延についての勧告であるが、そこでは音声遅延量が往復 300msec を超えると話に支障をきたすことが報告されている [23]。

2.3 非言語情報

コミュニケーション時にメッセージを伝達するチャネルは数多くあるが[12]，それらを言語情報と非言語情報の観点から分類する方法が主流である[14, 33, 49, 56]．

人間特有の言語情報を利用して行われるコミュニケーションは言語コミュニケーションと呼ばれる．例えば，手紙のように文字を用いたコミュニケーションは言語コミュニケーションの典型であるし，直接相手と対面して行う会話の中にも言語コミュニケーションの要素は多く含まれている．

一方，非言語情報を利用して行われるコミュニケーションは非言語コミュニケーションと呼ばれる．例えば，うなずき，瞬き，注目している方向，視線，しぐさ，容姿，ジェスチャー，声の強弱，声のピッチ等を用いたコミュニケーションは非言語コミュニケーションである．

2.3.1 非言語情報の役割

非言語コミュニケーションが社会的相互作用の中で果たす役割は大きく，中でも次の機能が重要であると Patterson は述べている[47]．

- (1) 情報提供
顔の表情等により，相手に伝えたい情報が補完される．
- (2) 相互作用の調整
頷きや姿勢等により，会話を続けるか，話題を変えるかといったように，相手との相互作用の調整が行われる．
- (3) 親密さ表出機能
対人距離を小さくすることで，相手との親密度の高さを示すことができる．
- (4) 社会的統制機能
声を大きくすることで，発話者が発言に自信を持っていることを他者に示せる．
- (5) サービスと仕事の機能
身体接触等により，相手との親密度を上げることができる．

このように，非言語情報はコミュニケーションの中で重要な役割を果たしており，人間のコミュニケーションの約 65 % が非言語情報によって成立しているとの報告[1]もある．

表 2.1: 非言語情報の分類 [14, 47]

非言語情報	具体例
対人距離	空間的配置, 他者間距離
体の動き	体の向き, 体の動き, ジェスチャー, 姿勢
表情	微笑み, 顔をしかめる
視線	誰の方向を向いているか, 凝視しているか
接触	触れる, 抱き合う
準言語	話の間, 声の大きさ, 流暢さ
嗅覚作用	他者からの香り
人工物	化粧, 服装, 装飾品

2.3.2 非言語情報の分類

非言語情報に関する研究事例は数多く, その分類法も多様である [13, 14, 34, 44, 47, 67]. 表 2.1 に示すのは, Patterson による分類 [47] に原岡らが解釈を加えたもの [14] である. 以降, 分類分けされたそれぞれの非言語情報について説明する.

- 対人距離

原岡らの定義によると, 対人距離とは相互作用において人と人とがとる物理的距離であり [14], コミュニケーションにおいて重要な役割を果たしている. 例えば, 近い距離で話す 2 人は親密である場合が多く, とりわけ男女の関係においては顕著である.

また, 会議の際の席順もコミュニケーションに大きな影響を与える. 特に日本では, 出入り口から一番遠くに座っている人が一番発言権がある場合が多く, 座る位置次第で発言頻度・内容が変わることも少なくない. 席を自由に選ぶ際に仲が良い者同士が近くに座ることも, 対人距離がコミュニケーションに大きな影響を与えている例の 1 つである. Kirstie Hawkey らは, 対面コラボレーションでの距離の影響を調査した [15].

- 体の動き

人がコミュニケーションを行う際に意識的・無意識に行うジェスチャー等の体の動きは, 非言語情報の中でも重要な役割を果たしている [44]. 堂々とした姿勢で大きなジェスチャーを交えて発言する話者は自信があるように見えるし, 悪い姿勢でうつむきながら発言する話者は頼りなく見える.

コミュニケーションにおいて, 相手の体の動きから相手が自分の話に関心を持っているかを認識できる. たとえば話している最中に, 相手が全く違う方向を向いていたら, 自分の話している内容に関心を持っていないと認識できる. 一方, 相手が体

を自分に向け、身を乗り出して聞く等の体勢をとれば、それは自分の話に関心を抱いている可能性が高いことを認識できる [12]。このように体の動きは、人間のコミュニケーションにおいて重要な役割を果たす。

- 表情

表情は、人間の感情等を認識する上で重要な役割を担う非言語情報である。例えば、話者は自分の話に関心を持たれているかどうか、聞き手の表情から判断する 경우가少なくない。表情の中でも、瞬きの頻度は相手に関心を持っているかどうかということと深い関係があると報告されている [57, 58, 66]。

- 視線

“目は口程に物を言う”という言葉があるとおり、相手がどの方向を見ているか、どのような目つきで見ているか等の視線情報も、コミュニケーションでは重要な役割を果たしている。話者は聞き手の視線を確認しながら話す場合が多い。話者は自分に視線が向けられていれば聞き手が興味を持っていると判断できるし、逆に視線を向けてもらえなければ聞き手が興味を持っていないと推測できる。

- 接触

接触は、触れる・抱き合う等の非言語情報である。人はコミュニケーションをする際に、相手の肩をたたいたり頭をなでたりすることで、言語情報だけでは表現できない親密度を表現することがある。また、スポーツや舞踊のレッスンでは、接触を通じてスキルの伝達も行われる。

- 準言語

話の間・声の大きさ等の非言語情報は準言語と呼ばれている。これも重要な非言語情報の1つであり、準言語の使い方次第で情報伝達の効率や伝わり方が大きく異なる。例えば、大きな声で淀み無く発言する話者が自信を持っているように見える現象は、話の内容そのものよりも準言語によるところが大きい。

- 嗅覚作用

嗅覚作用とは、非言語情報の1つである香り等を通して嗅覚を刺激する作用のことであり、コミュニケーションにおいて重要な役割を果たしている。例えば、心地よい香りの香水をつけている人は好意を持たれやすいし、口臭のきつい人は敬遠されることが多い。

- 人工物

人工物とは、化粧・服装・装飾品等のことであり、コミュニケーションにおいて非言語情報を伝えている。人は身だしなみである程度相手の人格・社会的地位・嗜好等を判断することが多く、人工物がコミュニケーションにおいて重要な役割を果たしていることが分かる。

- 予備動作

Vargas は、人が対面したコミュニケーションをする際に、発話権を獲得するために様々な非言語情報を活用していることを明らかにした [67]。この非言語情報には、相手に分かるように強く頷くことや、注意を引くように相槌を打つこと、組んだ足を下ろす、腕組みをほどく、身体を話者の方へ向ける、前へ乗り出す、手を挙げる動作などがあるとしている。

2.4 対面コミュニケーションの支援に関する先行研究

我々は日常的に他者とコミュニケーションを行う。よく知り、慣れている相手とのコミュニケーションは問題なく行えるが、初対面の相手や、慣れていない相手とのコミュニケーションは必ずしも円滑に行えるとは限らない。これをサポートするために、モチベーションを高める方法 [26] や、適切な話題を提供することで会話を盛り上げることを目指した研究がある。

2.4.1 適切な話題の提供支援

本節では、対面コミュニケーションを行う際に、その場面や相手に応じて適切な話題を提供することで会話を盛り上げることを目指した先行研究を紹介する。

- SCACS

SCACS(a Social Context-Aware Communications System) は、研究者達が集う場における対面コミュニケーションにおいて、各研究者の所属学会や共著者情報（誰と誰が同じ論文で共著の関係にあるか）を提示することで適切な話題を選択することができ、コミュニケーションが円滑に進むことを狙ったシステムである。このシステムは主に下記コンポーネントから構成されている。

- SNS Server:
文献検索サービスを利用して収集した各研究者の共著情報が格納されている。SNS Server 内に格納された情報は API を通じて後述の Portable Computer から取得できる。
- Environmental Sensor:
ユーザの向きや相手との位置関係等を測定するための機器であり、GPS やジャイロセンサがこれにあたる。取得した情報は Portable Computer に送信される。
- Portable Computer:
EnvironmentalSensor から得られた情報に基づいて、コミュニケーション相手に関する情報を SNS Server から取得し、後述の Display Device 上に適切な形態で表示する。

- Display Device:

Portable Computer から受信したコミュニケーション相手の背景情報を提示するためのデバイスであり，HMD (Head Mounted Display) や PDA，携帯電話等がこれにあたる．

実際に学会においてこのシステムを利用したところ，背景情報を把握した上で行うコミュニケーションは相手の情報が全く分からない場合よりも円滑に進行することが報告されている．

- MeetBall

MeetBall は，話題にしている対象の映像をテーブルトップへ提示することで発話を活性化させるコミュニケーション支援システムである [37]．

このシステムは情報処理用のコンピュータや映像提示用のプロジェクタを内包した球状のデバイスであり，コミュニケーションが行われるテーブルの上方から吊り下げる形態で利用する．コミュニケーション参加者の発話内容から音声処理技術によって単語を抽出し，その単語に適合する画像をテーブルトップに映写して発話内容を視覚化することで会話の活性化を実現している．例えば，一部の参加者しか知らない事物が言及された場合でも，それが MeetBall によって映像化されることで全員が視認する事ができ，会話が中座してしまったり一部の参加者だけで会話が盛り上がったってしまうことを防いでいる．その他，参加者ごとの発話の偏りをシステムが検知し，発話の少ない参加者へ発話を促すファシリテーション機能も有する．

2.4.2 人とエージェントとの会話

ここでいうエージェントとは，ユーザとシステムの間に入り，システムの操作や，ユーザとシステムのインタラクションを支援するロボットやアバタなどを指す．そのエージェントの振る舞い方を人の行うそれに近づけることが主な研究の方向性である．Chidambaram らはロボットと人とコミュニケーションする際に，ロボットがジェスチャを使うことで受け手となる人へメッセージが伝わりやすくなることを明らかにした [5]．また Mutlu らはロボットの自然な視線行動を実現するための方法について研究した [36]．さらに，人と機械の会話を自然に感じさせるために発話前の非言語情報を表現する機能をエージェントシステムに組み込んだ試みなども存在する [4, 29]．

2.5 遠隔コミュニケーションの支援に関する先行研究

遠隔コミュニケーションシステムは電話に始まり，やがて音声だけでなく映像も通信する映像コミュニケーションシステムが開発された．さらに参加者の映像と音声のみならず，様々な電子データを送受信できるようになったため，会話機能だけではなく，遠隔地間の協調作業を支援する機能を有するものが多く存在する [17, 40, 42, 72]．これらは遠隔コラ

ボレーションシステムと呼ばれ、遠隔地にいる聴衆とのインタラクティブなプレゼンテーションを支援するもの [20]、デスクトップ上での作業に限定して機能を絞り、軽い処理でインタラクションが行えるもの [70]、オンライン教育に焦点を当てたもの [16, 75]、セカンドライフ内でのコラボレーションを支援するもの [30] など、幅広く研究が行われている。

本研究では遠隔での会話機能に対象を絞っている。この基本的な機能に関して、1990年代に Sellen がビデオ会議システムにおけるコミュニケーションの弊害を指摘して [55] から未だに解決されたとはいえず、研究は継続されている。その研究の方向性は主に次の2つである。1つは映像、音声の質を高めて、対面しているのと同レベルの環境に近づける方向性で、専用の機材を用いて大規模なシステムを構築することが多い。もう1つはインターネットに接続したノートパソコンなどから参加可能な小規模のシステムで、参加者をその化身となるアバタで代用し、アバタの外観や行動を実際の人のそれへと近づける方向性である。

2.5.1 テレプレゼンスを目指した遠隔コミュニケーションシステム

遠隔地の参加者があたかも目の前にいるように感じるテレプレゼンスを目指したシステムが多く研究されている [6, 27]。

- Hydra

Hydra はテーブルを囲んでいるような状況を再現する遠隔会議システムであり、トロント大学で開発された [3]。

カメラ・モニタ・スピーカをそれぞれ1台ずつ使用し、1つの画面中に同時に全参加者を映し出す一般的な在席会議システムの方法は、顔の向きや視線等で会話の流れを制御できないため次のような問題がある。

- 個々の参加者に注意を注ぐことが難しい。
- 自分が誰かに注目されていてもそれに気付くことが難しい。
- 2人以上の同時の発話を聞き分けることが難しい。
- 2組以上に別れて同時に会話をするのが難しい。
- 内緒話をするのが難しい。

上記を改善するために、Hydra では小型のカメラ・モニタ・スピーカが一体化されたユニットが遠隔地にいる参加者に1式ずつ割り当てられ、これらがあたかも対面環境に参加者が存在しているかのように配置されている。このように、人がいる位置（いると想定する位置）にモニタを配置する発想は Hydra 以前からあり、Bell Northern Research 等で以前から研究されている。ただし、Hydra はユニットが小型であり、比較的省スペースな環境にて遠隔会議を行えるという特徴がある。

しかし，Hydra と1つの画面中に同時に全参加者を映し出す一般的な在席会議システムの比較実験によると，並列な会話が可能になったこと，会話を追うのが楽になったことは確認できたが，その他には大きな差異が認められなかったと報告されている．これは，画面が小さすぎたために視線を追う効果が減退されたことが一因とされている．

- cAR/Pe!

Hydra が現実空間中に各ユーザのモニタを配置したのに対し，cAR/Pe!は仮想空間中において円卓を囲むように各ユーザを配置している [48]．各ユーザはビデオ映像のライブストリーミングで表示されているため，表情等の情報を欠落させずに伝え合うことが可能である．さらに，各ユーザは自分の映像が表示されているパネルを回転させられるため，誰が誰の方向を向いているのか他者が理解でき，円滑にコミュニケーションを進行できるというメリットがある．

- MAJIC (Multi-Attendant Joint Interface for Collaboration)

臨場感が無い，視線の一致を行えない等の既存システムの問題点を克服した遠隔会議システムである [43]．MAJIC は次のコンセプトに基づいて設計されている．

- － 等身大の相手画像との視線一致
等身大の表示を行うことで，より細かな仕草の伝達が可能になるだけでなく，身体動作を使ったコミュニケーションも活発になることが期待される．
- － 1枚の湾曲スクリーンにシームレスに投影
一人一人を四角い枠に閉じ込めることなく，全参加者を1枚のスクリーン上に投影し，後ろに連続した1枚の背景を合成することで人と人之間をシームレスにできる．このようにシームレス映像を映し出すことで，臨場感・一体感の向上が期待できる．
- － デスクトップの作業領域
MAJIC では，画面と利用者の間の空間に連続したデスクトップの作業領域を設けることができる．これにより，対面環境と同様に会議に必要な共同作業領域を設けることが可能になる．また，卓上の作業空間は「視線を自然に外せる自由」として活用できるため，常に相手の顔を見ることを強いられず，より自然で人間的な会議空間が実現できる．

MAJIC では建装材として市販されているコントラビジョンを利用している．これは，透明なシートに特殊な技術を用いてドット等のパターンを印刷した物で，シートの表裏を全く異なったデザインにすることが可能である．表が白，裏が黒の細かなドットが印刷されたコントラビジョンを利用することで，表からはスクリーンになり，裏からは透けて見えるような効果を得られる．このスクリーン上に多地点にいる相手の画像を等身大で投影し，相手画像の顔の真後ろにカメラを配置してこち

らの様子をスクリーン越しに撮影すると、相手の目を見つめる行為がその後ろにあるカメラを見つめることになり、視線の不一致が起こらない。

スクリーンは 1.2m の半径の円周に沿うように湾曲されて設置され、そこにそれぞれの相手画像が対面環境にいるかのような位置関係で投影される。この円の中心に利用者が座ると、まるで他の参加者と同じ部屋にいるような一体感・臨場感を得ることができる。また、投影画像の大きさと位置を変化させることで、仮想会議室内の座席配置を自由に変えることもできる。

- Montage

分散コミュニケーション環境において glance(一瞥) をサポートしたデスクトップビデオ会議システムである [63]。この glance が提供されることにより、分散環境にいるユーザ達は周りの状態に応じて話しかけることが可能になる。この研究では、相手の所在確認や相手と話す機会を交渉することは pre-interaction と定義されており、これが発生する状況のモデルとして下記の 3 つが挙げられている。

- telephone モデル

telephone モデルは、話をしようとしても相手が不在であったり、その場合に誰かが代理をしなければいけなかったりする等、生産性は必ずしも高くはないが、プライバシーに関しては完全に守られている。

- overview モデル

overview モデルとは、各ユーザの状況が常に撮影されて、全員が互いの状況を把握できるモデルである。このモデルでは、相手の様子が分かる反面、自分が常に監視されているのでプライバシー保護の観点では問題がある。特定の状況においては有効であるが、それ以外の場合はユーザはカメラに映らないように隠れてしまうことが多い。

- hallways モデル

人が誰かと話すためにオフィスや廊下を歩き回る様子をモデル化したものである。このモデルでは、プライバシーの保護と制御が行いやすく、コミュニケーションをするための適切な機会を見つけやすい。

Montage は hallways モデルに基づいて構築されている。ネットワークを通して相手と映像・音声の通信を行うことができるが、その際はいきなり映像・音声が接続されるのではなく、始め映像のみが接続されて (glance)、話しかける方はこの映像を見て話すか話さないか判断できるようになっている。そして、話すと判断した場合、次に音声も接続される。映像が現れる際も、唐突に現れるのではなく、最初に足音がして、それに伴って画面がフェードインする工夫が為されている。また、相手が不在等で話せない場合への対策として、Montage ではオンラインカレンダー・Stickup note (スクリーンに注意書きを残しておく機能)・eメール等のツールが提供されており、同期から非同期への通信にもある程度配慮がされている。

9人からなるグループが9週間に渡り行った評価実験によると、大部分の glance はすぐには双方向のコミュニケーションには発展しなかった。これは、Montage が簡単にコンタクトをとれることを支援しているだけでなく、コンタクトに至る pre-interaction を支援できていることを示している。

Montage は、ユーザがコミュニケーションを行うための適切な機会を発見するのに有効な映像的きっかけを提供している。つまり、Montage は音声だけでなく映像が双方向のコミュニケーションにおいて有効であることを実証しただけでなく、映像がどのように pre-interaction を支援するべきかについても探求したと言える。コミュニケーションが始まる前の交渉を自然な形で行うことを可能にしたこのアプローチは従来のデスクトップビデオ会議システムの映像の使用法とは明らかに異なるものである。

2.5.2 アバタを用いるシステム

遠隔コミュニケーションシステムにおいて、遠隔地にいる相手をアバタで置き換えることによって会話を円滑にすることを試みている例も多数存在する。

- 石井らのシステム

石井らはアバタを介した遠隔コミュニケーションシステムにおいて、次に発話することを促す視線を表現する手法を提案した [21]。発話音声の音圧や周波数の変化から話者交替のタイミングを判別する。そのタイミングで話者や他の参加者のアバタが聞き手側へ視線を向けるように見せることで、次に発話することを促す。しかしこの手法では、音声を基に視線動作を疑似的に作り出しており、実際の参加者の動作は反映されない。

- 川島らのシステム

川島らは2者間の遠隔コミュニケーションにおいて、フィラー（間を埋める音声）を、映像によって表現した [24]。しかし参加者は2人に限定され、これも実際の参加者の動作は反映されない。

- PAW2

インターネット上に構築された仮想空間中にユーザがアバタとして参加するタイプのコミュニケーションシステムである [31]。操作は基本的にマウスとキーボードを用いて行われるが、アバタの表情を変化させるインタフェースが用意されており、各表情のボタンをクリックすることでユーザは任意のタイミングでアバタの表情を変化させられる。

- InterActor

渡辺らは、顔きや身振り等を「コミュニケーションの引き込み動作」と定義して、これを音声から自動生成する音声駆動型身体引き込みキャラクタ InterActor というシ

ステムを提案した [69]。InterActor は音圧の ON / OFF 情報に基づき、MA モデルを用いることで呼気段落の切れ目を予測する。予測した呼気段落の切れ目にアバタの頷き動作を生成することで、発話を活性化させる。1つの予測モデルを用いて、話の聞き手としてのジェスチャと、語り手としてのジェスチャの両方を生成することができる。

2.5.3 その他のとりくみ

アバタを介してコミュニケーションを行うものの中にも、大規模な装置を用いてアバタの動きをよりリアルにしようとする試みもある [19, 22, 38, 68]。VirtualActor はユーザが頭頂部、背中、両腕に装着した4個の磁気センサーの角度・位置情報を利用してユーザの頷き、身振り等の被言語情報を表現している。また、データグローブを用いて手指の動きを入力することでアバタのジェスチャをより豊かにする試みも為されている [45]。

遠隔コミュニケーションにおいても、人同士ではなく、人とエージェントとの遠隔コミュニケーションを支援するための研究も存在する。Burke らは、人とロボットが遠隔で会話する際の映像情報の効果の調査を行った [2]。

また、テキストチャットを組み合わせた例も存在する。Czerwinski らは、インフォーマルコミュニケーションにおいてチャットとリッチなメディアを使ったコミュニケーションとの比較を行った [8]。インフォーマルコミュニケーションでは音声コミュニケーションよりもテキストチャットが便利であると答えたユーザが多かった。その理由として下記の項目が挙げた。

- ・非同期のコミュニケーションができる
- ・話しかけることにより相手を妨げる度合いが低い
- ・周囲の雑音の問題を受けない
- ・大きな規模のグループでも会話しやすい
- ・母国語以外のコミュニケーションをしやすい

また、Web カメラに映った映像をテキストチャットと組み合わせることにより、下記の点でコミュニケーションを拡張できている。

- ・相手の存在感がよく伝わる
- ・相手の状態がよく伝わる
- ・感情的な反応がよく分かる

以上のように、テレプレゼンスの実現を目指した研究開発は盛んであるが、可搬性の高い Web 会議システムなどにおいては、意図的参加者の化身となるアバタを用いて、アバタに参加者の実際の動きを反映させるか、動きを仮想的に作り出すことが多い。しかし参加者の実際の動きを高精細に認識して再現することは困難であり、仮想的に作り出した動きは、参加者の意思を反映したものにならない恐れがある。このため本研究は、Web 会

議において映像と音声を共有する他に，会話を円滑にするために必要な参加者の非言語情報を1つ検出して伝達することで，円滑な話者交替を実現させることを目指す．

第3章 音声遅延と発話衝突確率および精神的ストレスの関係

3.1 遠隔コミュニケーションにおける話者交替の弊害

自席のデスクトップパソコンやノートパソコンから手軽に遠隔会議を行うには Web 会議システムが適している。本研究が対象としている創造会議は、話者交替が多く起こる会議である。そして話者交替の失敗は発話衝突と沈黙であると言われている。Web 会議程度のクオリティでは、映像から読み取れる情報は少なく、非言語情報は伝わりづらい。今日のベストエフォートの通信環境では、音声を優先させた場合、解像度やフレームレートが下がってしまうことが多い。徳らは1つのディスプレイ上で目状に表示される参加者映像は小さく、各々の様子をよく見ることができないと指摘した [65]。これらのことから、参加者は Web 会議システムを用いて会話をする際、主に音声情報に頼って会話していると考えられる。

音声情報の伝達を阻害する要因には遅延、サンプリングレート、通信帯域、エコーなどが挙げられるが、遠隔で会議を行う場合に、最も大きく、なくすことができない要因は音声遅延である。音声品質の中でも音声遅延は、発話衝突を引き起こす大きな要因であると考えられる。音声遅延による、会話への影響を調査した研究があるが、音声遅延が引き起こす発話衝突の度合いを定量的に調査した例は存在しない。鎧沢らは、2 地点間の映像コミュニケーションを行う際に、交互に数字を読み上げるような極端に速い話者交替を必要とする会話は、音声遅延量が増加するほど、会話が妨害される度合いも増すが、特別速い話者交替を必要としない会話であれば往復 500msec まで音声遅延が生じて問題はないと報告した [73]。ITU-T 勧告 G.114 では、往復 300msec を超えると話に支障をきたすと報告されている [23]。そこで、音声遅延量と発話衝突確率の関係や、発話衝突が参加者に与える受ける精神的なストレスの度合いを調査する。

3.2 音声遅延による発話衝突のメカニズム

本節では音声遅延による発話衝突のメカニズムを議論する。まず、図 3.1 のような話者交替時の認知・行動モデルを考えた。ある参加者が発話欲求を持ったとき、意識的か無意識的に関わらず何らかの動作を表出する。そして他の参加者の様子を観察し、他の参加者が発話しようとしているか、もしくは既に発話しているかを知る。その情報を基に、自身が発話するのか、それとも沈黙して他の参加者へ発話を譲るのかを判断する。そして実際に発話、もしくは沈黙するといった行動に移る。対面コミュニケーションではこのステップが滞りなく流れ、円滑に話者交替が行われている。

続いて音声遅延のある環境で発話衝突の起こるメカニズムについて、話者交替のモデルに基づいて説明する。ここでは分かりやすくするため、音声の遅延のみに焦点を絞って説明する。2 人の参加者 A と参加者 B が会話をしているとき、話者交替時の認知・行動モデルは図 3.2 のように遅延の影響を受ける。参加者 A に発話欲求が生じたとき、他の参加者が発話しているかどうかを観察する。他の参加者が発話せず、自身が発話しようと判断したとき、実際に発話にいたる。発話した音声は、他の参加者のスピーカで、音声遅延時間分だけ遅れて再生される。これと同じモデルが、参加者 B にも適用される。図 3.2 の例で

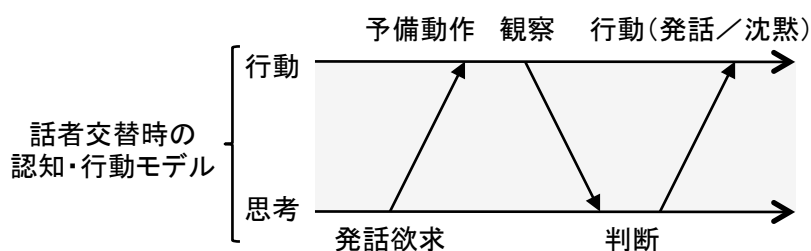


図 3.1: 話者交替時の認知・行動モデル

は、参加者 B が観察している時刻には、参加者 A は実際には発話を開始している。音声遅延がない環境であれば、これに参加者 B は気づき、発話を中止することができる。しかし、上記例では音声遅延があるため、参加者 B が観察している間は、参加者 B には参加者 A の音声がかえって聞こえていない。そのため参加者は発話可能であると判断し、発話を開始する。結果的に参加者 B の発話した音声は、参加者 A のそれと衝突してしまう。音声遅延量が増加するほど、この問題は大きくなり、発話衝突が生じる確率（発話衝突確率）が増加する。また、音声遅延のある環境で発話衝突が生じた場合、その状態から復旧する（いずれかの参加者が発話権を譲り、もう一方がそれを獲得する）ことが、下記の理由により、困難になると考えられる。

1 つは実際の発話開始順序と、参加者の体感する発話開始順序が食い違うことである。図 3.2 の例で、参加者 B は参加者 A より後に発話開始しているが、参加者 A の音声を聞き始めるのは自身が発話した時刻よりも後である。このため参加者 A、B ともに自分が先に発話開始したと感じ、発話衝突後も両者が発話を継続しようとしてしまう可能性がある。

もう 1 つは発話衝突の状態から復旧するための行動も、音声遅延量だけ遅れて相手へ伝わることである。すなわち発話衝突を認知し、判断した結果行った沈黙もしくは発話継続という行動も、音声遅延量だけ遅れることになる。前述の例のように、2 人で会話をする場合、話者交替適格場において、次に発話する可能性のある参加者は 2 人であり、現話者が発話を継続するか、聞き手が発話を開始するかのどちらかである。しかし、参加者が 3 人に増えると、TRP において、次に発話する可能性のある参加者は 3 人に増える。その場合、図 3.2 のモデルはさらに複雑化する。本研究は 3 拠点以上をつないだ Web 会議を対象としている。そのため今回は、その最小地点数である 3 地点を結ぶ会議を行う際の音声遅延による影響を実験により調査した。

3.3 実験手順

3 地点間で音声をつなぎ、司会者を設けずにアイデアを創出する創造会議を実施した。音声遅延量と発話の衝突確率、精神的ストレス、創出されたアイデアの数の関係を調べた。この創出されたアイデア数は、既存研究では発想支援システムの有効性を測る指標

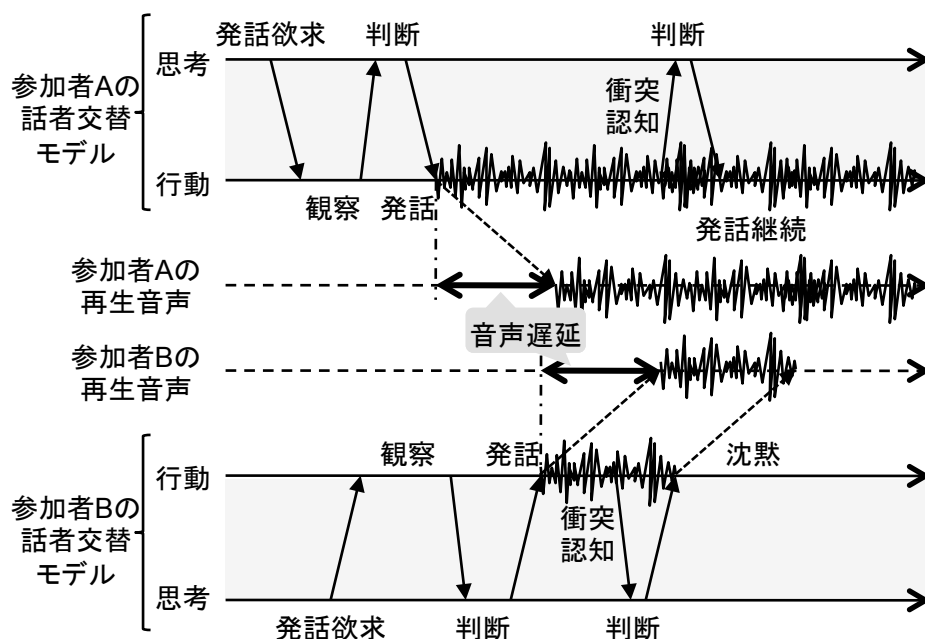


図 3.2: 音声遅延による発話衝突への影響

などに用いられている [74] が、本実験では会話の効率性を測る指標として用いた。音声遅延の影響により発話回数が減少するか、発話衝突して言い直す回数が増加することで、会話が円滑に進まなくなることが考えられる。複数の参加者で行う Web 会議では 1 人の映像領域が小さくなり、帯域の保証されないネットワーク上で実施されるため、音声品質をある程度高く保とうとし、映像の解像度とフレームレートが低くなることが多く、映像は活用されていない [65]。そのため映像の影響は小さく、音声のみを使った環境で実験をすれば、映像のある Web 会議にも適応可能であると考えられる。鎧沢らの調査 [73] において、音声会議環境と、映像を含む会議環境で、音声遅延の妨害度合いを調査した際に、2 つの会議環境の間で同じ傾向が見られたことが報告されている。大型ディスプレイを用い、帯域の保証されたネットワーク上で通信を行うテレビ会議などでは映像自体の遅延、映像と音声の相対遅延などが影響し、本実験とは傾向が異なる可能性がある。

図 3.3 に、音声会議環境を示す。被験者はそれぞれ、個別の防音室に入り、ヘッドセットを装着した。各被験者の装着するヘッドセットのマイクから入力された音声はディレイマシン（遅延装置）に送られた。ディレイマシンは音声遅延を 0 から 1 秒まで調整することができた。ディレイマシンから出力された音声は、入力元以外の被験者が装着するヘッドセットのスピーカへ送られた。つまり、それぞれの被験者は自分以外の被験者の、ディレイマシンを 1 度経由した音声を、ヘッドセットを通して聞くことができた。

20 代から 30 代の男女 30 人の被験者を 3 人一組の 10 グループに分け、決められた時間内に、できるだけ多くのアイデアを考え出す会議を行わせた。グループを構成する 3 人は、同年代の友人であることを条件とした。音声遅延量を 0msec、200msec、400msec、600msec、

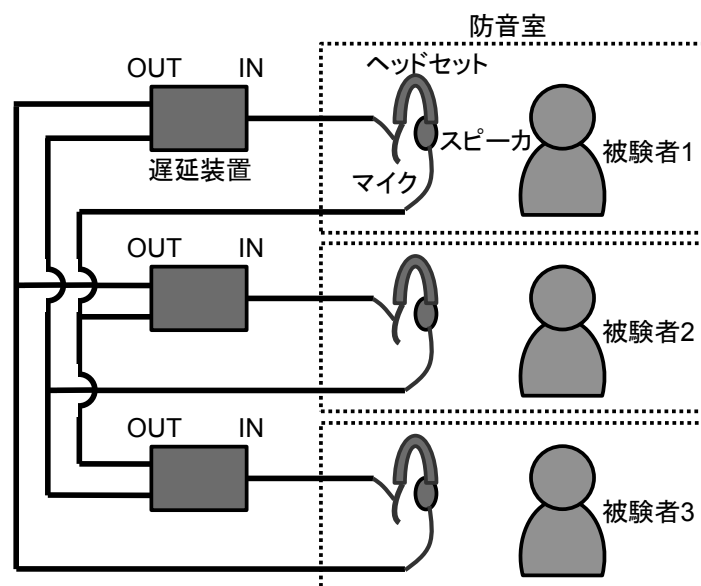


図 3.3: 音声遅延会議システム

800msec, 1000msec の 6 種類用意した。この範囲に設定した理由は、遅延が全くない状態から、実際のインターネットを通じた会議を行う際に発生しうる値を十分にカバーする範囲で実験するためである。各グループは 6 種類すべての遅延環境で会議をし、その順序はグループごとにランダムに入れ替えた。1 つの条件ごとの会議時間は 7 分間とした。

会議のテーマは「携帯電話の新機能を考える」とし、6 つの条件の間で継続させた。同じテーマで会議を行う場合に時間経過とともに発話回数が減少するなど、順序効果による影響が懸念された。しかし 6 つの条件ごとにテーマを変えると、それぞれのテーマに関する被験者の前提知識や興味の度合いなどが大きく影響してしまう。このため同じテーマを継続して用い、体験する遅延時間の順序を被験者グループ間でランダムに入れ替えることによりこのような順序効果を解消する実験計画を採用した。3 人のうち 1 人を書記役とし、会議に参加しながらグループで考えたアイデアを記録させた。各条件で会議を実施した後、質問紙による主観評価を実施した。主観評価項目は下記の通りであった。5 段階評価で 1 は「まったく感じなかった」、5 は「強くそう感じた」とし、項目ごとに当てはまる得点をつけさせた。

- 音声遅延をストレスに感じる
- 誰が話しているか分からないことをストレスに感じる
- 誰が話し始めそうかわからないことをストレスに感じる
- いつ話し始めていいかわからないことをストレスに感じる
- 話し始めが他の人とぶつかることをストレスに感じる

- 他の参加者が自分の話を聞いているか分からない
- 沈黙が多い
- 話に割り込めないことをストレスに感じる
- 盛り上がらない
- 相手の存在感が薄い

会議中の被験者の映像，音声をビデオに記録し，分析した．ビデオを解析し，すべてのグループの条件ごとの発話回数，発話衝突回数を計測した．相槌や笑い以外の意味を持つ音声入力を発話とみなした．このとき，衝突などによってターンを取得できなかった発話も回数に数えた．アイデア創出数に関しては，被験者の記録したアイデアの数を数えた．このときにアイデアの質などは考慮せず，単純に創出された数のみを計測した．

3.4 実験結果

3.4.1 発話衝突確率

音声遅延量ごとの，各グループの7分間での発話回数の平均を図3.4に示す．二元配置分散分析を行った結果，遅延量による発話回数の差異は認められなかった．どの条件でも1分あたり10回程度の発話があり，1つの発話は4秒前後であった．どのグループでも3人の被験者は，書記役であるか否かに関わらず，偏りなく発話していた．本実験で実施した司会者のいないアイデア創造会議は頻繁に話者交替をとまなうことが多いといわれている[18]．本実験結果からは，発話の長さや話者交替の間隔への音声遅延による影響は見られなかった．

図3.5に，発話した回数を母数とし，そのうち発話衝突が起こった回数の占める割合を発話衝突確率として示す．音声遅延0msecと200msec条件の間で発話衝突確率はほぼ変わらず，音声遅延量600msec条件で最大の値をとる結果となった．二元配置分散分析（群間自由度5，群内自由度9）を行った結果，F値2.56，P値は0.04であり，有意水準5%で音声遅延量により発話衝突確率に差があるという結果が得られた．Tukeyの方法により多重比較を行ったところ，有意水準5%のとき，音声遅延量0msecと600msec，200msecと600msecのときの発話衝突確率の平均差はそれぞれ5.42，6.08となり，これらの条件間で発話衝突確率に有意な差が認められた．

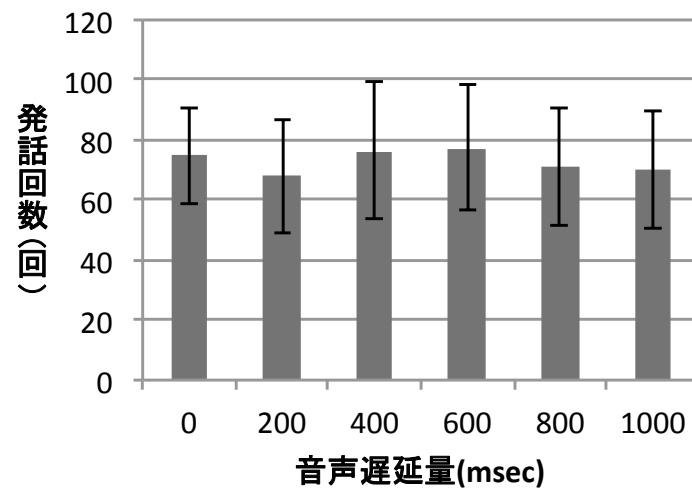


図 3.4: 7 分間での平均発話回数の比較

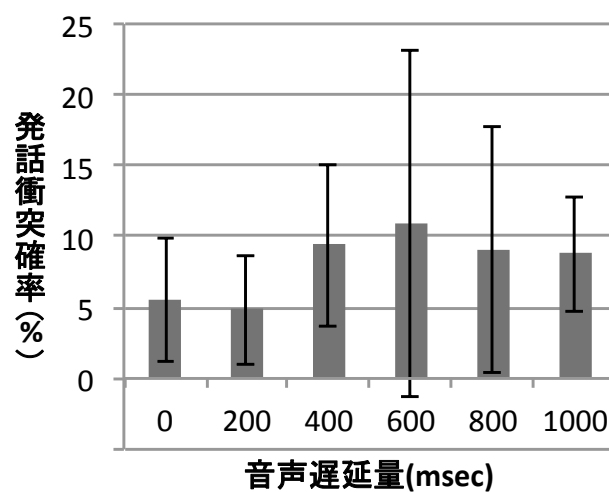


図 3.5: 発話衝突確率

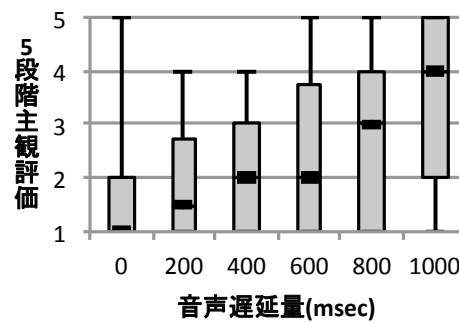


図 3.6: 質問紙評価項目「音声遅延をストレスに感じる」の結果

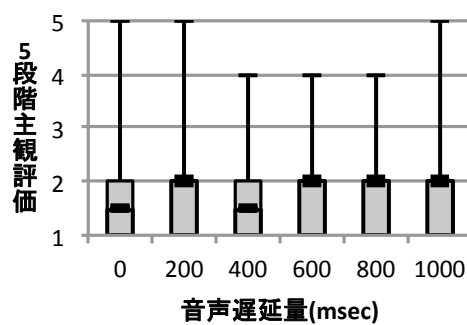


図 3.7: 質問紙評価項目「誰が話しているか分からないことをストレスに感じる」の結果

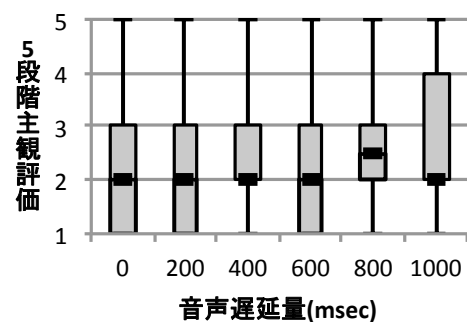


図 3.8: 質問紙評価項目「誰が話し始めそうか分からないことをストレスに感じる」の結果

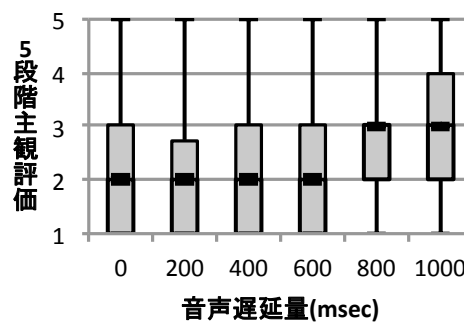


図 3.9: 質問紙評価項目「いつ話し始めていいかわからないことをストレスに感じる」の結果

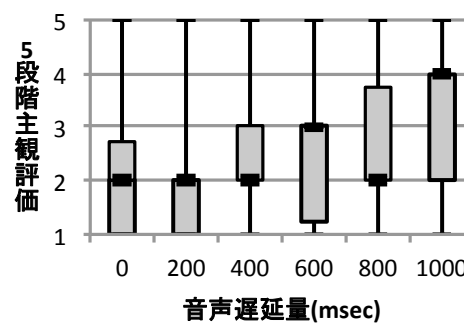


図 3.10: 質問紙評価項目「話し始めが他の人とぶつかることをストレスに感じる」の結果

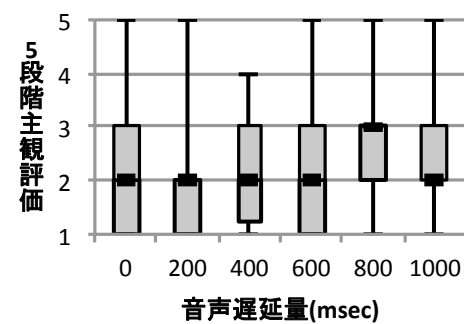


図 3.11: 質問紙評価項目「他の参加者が自分の話を聞いているかわからない」の結果

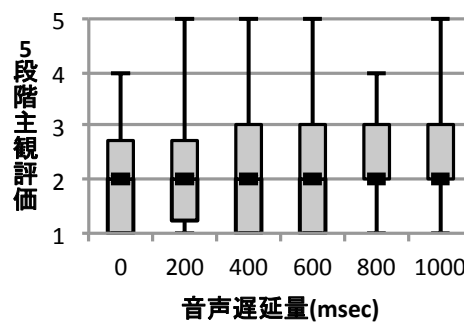


図 3.12: 質問紙評価項目「沈黙が多い」の結果

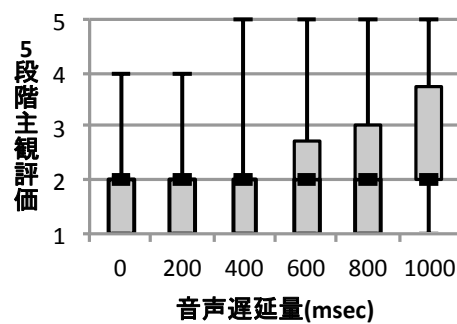


図 3.13: 質問紙評価項目「話に割り込めないことをストレスに感じる」の結果

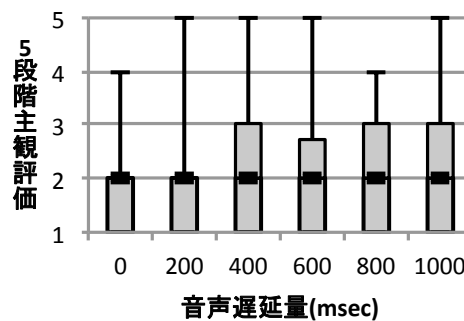


図 3.14: 質問紙評価項目「盛り上がらない」の結果

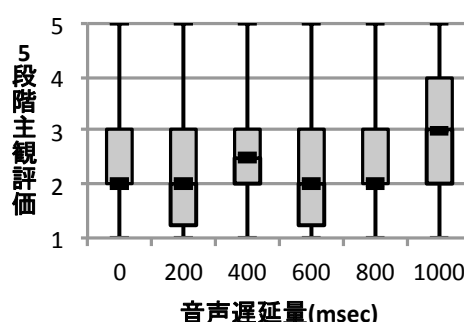


図 3.15: 質問紙評価項目「相手の存在感が薄い」の結果

3.4.2 精神的ストレス

質問紙による主観評価の結果を図 3.6 から図 3.15 に示す。各図中の太線は中央値を示す。またそれぞれの結果について Friedman の検定を実施した結果を表 3.1 に示す。音声遅延量の変化が主観評価の結果に影響するという結果が得られたのは「音声遅延をストレスに感じる」「話し始めが他の人とぶつかることをストレスに感じる」の 2 項目である。この 2 項目について、Bonferroni-Holm 法により有意水準の調整を行った後に Wilcoxon の符号順位検定を用いた多重検定を行い、音声遅延条件ごとの組み合わせで比較を行った結果を表 3.2, 3.3 に示す。「なし」は統計的有意差が認められなかった項目、「*」は有意水準 5 % で差が認められた項目、「**」は有意水準 1 % で差が認められた項目である。有意水準 1 % で、音声遅延量 800msec の条件では 0msec の条件よりも「音声遅延をストレスに感じる」、音声遅延量 1000msec の条件では 0msec, 200msec, 400msec, 600msec の各条件よりも「音声遅延をストレスに感じる」という結果が得られた。また、音声遅延量 1000msec の条件では 0msec, 200msec, 400msec の条件よりも有意水準 5 % で「話し始めが他の人とぶつかることをストレスに感じる」という結果が得られた。多重比較の結果、有意差の出た組み合わせは以上のとおりであったが、音声遅延量 0msec の条件と同程度にストレスを抑えられているのは 200msec の条件のみであった。音声遅延量が 400msec や 600msec の条件では発話が衝突することによるストレスを徐々に感じ始め、音声遅延量が 800msec を超えると音声遅延をストレスに顕著に感じるようになった。

3.4.3 会話の効率

条件ごとに、記録されたアイデアの数を集計し、7 分間で創出されたアイデアの数の 10 グループの平均値を表 3.4 にまとめる。創出されたアイデア数の音声遅延量による差は見られなかった。およそ 100 秒に 1 つ程度のアイデアが創出されたことが分かるが、今回のようなアイデア出しの活動は、数百ミリ秒単位の遅延量の違いによって影響を受けない

表 3.1: Friedman の検定結果

アンケート項目	自由度	カイ二乗値	有意確率 (P 値)	検定結果
音声遅延をストレスに感じる	5	29.50	0.00	**
誰が話しているか分からないことをストレスに感じる	5	2.69	0.75	n.s.
誰が話し始めそうか分からないことをストレスに感じる	5	2.34	0.80	n.s.
いつ話し始めていいか分からないことをストレスに感じる	5	9.81	0.08	n.s.
話し始めが他の人とぶつかることをストレスに感じる	5	12.35	0.03	*
他の参加者が自分の話しを聞いているか分からない	5	4.86	0.43	n.s.
沈黙が多い	5	7.14	0.21	n.s.
話に割り込めないことをストレスに感じる	5	6.13	0.29	n.s.
盛り上がらない	5	2.15	0.83	**
相手の存在感が薄い	5	3.00	0.70	n.s.

結果となった。3.4.1 節で述べたとおり、7 分間に 70 回程度の発話が観測されたことから、どの条件でも平均すると 16 回発話するうちに 1 つのアイデアが創出されたことが分かる。なお、今回の実験においてアイデアの数は会話の効率性を測るための指標として用いており、個々のアイデアの有益さの度合いは加味していない。

3.4.4 考察

音声遅延量が 200msec 以下の条件では、発話衝突確率も、主観評価の結果から得られた精神的ストレスも、音声遅延無し条件とほぼ同等であった。音声遅延量が増加するに従い、発話衝突確率と精神的ストレスの結果は各々異なる傾向を見せた。これらについて、以下に考察する。

3.4.4.1 発話衝突と精神的ストレスの関係

「音声遅延をストレスに感じる」と「話し始めが他の人とぶつかることをストレスに感じる」という主観評価項目において、音声遅延が上昇するほど、精神的ストレスが上昇するという結果が得られた。「音声遅延をストレスに感じる」という項目に関して、実験中の

表 3.2: 多重比較結果「音声遅延をストレスに感じる」

	0msec	200msec	400msec	600msec	800msec	1000msec
0msec	-	なし	なし	なし	**	**
200msec	-	-	なし	なし	なし	**
400msec	-	-	-	なし	なし	**
600msec	-	-	-	-	なし	**
800msec	-	-	-	-	-	なし
1000msec	-	-	-	-	-	-

表 3.3: 多重比較結果「話し始めが他の人とぶつかることをストレスに感じる」

	0msec	200msec	400msec	600msec	800msec	1000msec
0msec	-	なし	なし	なし	なし	*
200msec	-	-	なし	なし	なし	*
400msec	-	-	-	なし	なし	*
600msec	-	-	-	-	なし	なし
800msec	-	-	-	-	-	なし
1000msec	-	-	-	-	-	-

ビデオを観察したところ、音声遅延量が 800msec、1000msec の条件では、開始後すぐに音声遅延に気づき、「すごい遅延がある」と言及する場面が多く見られた。しかし音声遅延量が 600msec 以下の条件では、このようなことはほとんど見られなかった。800msec、1000msec の音声遅延がある場合、例えば呼びかけに対して応答する音声、期待する時間よりも遅れて聞こえることなどにより音声遅延に気がつくとともに、それがストレスにつながる可能性が考えられる。反対に、音声遅延量が 600msec 以下のときは、遅延が生じていることにすぐに気がつかなかった可能性がある。その理由として、今回のシステムでは 1) 自分の発話音声のエコーが聞こえないこと、2) 映像を含まないため、音声と映像の相対的なずれを感じることがないことが考えられる。「話し始めが他の人とぶつかることをストレスに感じる」という項目に関して、音声遅延 600msec の条件でわずかに高

表 3.4: アイデア創出数

音声遅延量 (msec)	0	200	400	600	800	1000
アイデア数	4.4	4.4	4.3	4.9	4.2	4.1

い値となったのは、発話衝突確率が最も高かったことと関係がある可能性がある。発話衝突確率が音声遅延 600msec の条件で最も高くなったことに対し、発話衝突による精神的ストレスは音声遅延 1000msec の条件で最も高い結果が得られた。この原因として以下が推測できる。

- (1) 音声遅延が大きい程、発話衝突に気づくまで、時間がかかること。
- (2) 音声遅延が大きい程、発話衝突した参加者が、互いに実際に発話開始した時刻を知ることができないため、話者交替適格場で発話を開始し、意図せず発話衝突が起こったのか、それとも話者交替適格場以後に発話開始し、自身の発話を妨害して発話権を獲得しようとしたのか分からない可能性が増加すること。
- (3) 音声遅延が大きい程、3.2 節 1) 2) で述べた要因が増大し、発話衝突から復旧することが困難になること。

3.4.4.2 音声遅延と発話衝突確率の関係

発話衝突確率は、音声遅延量が 600msec の条件で最も高くなり、800msec, 1000msec 条件ではそれよりわずかに低い値となった。前節で、音声遅延量 200msec から 600msec までの条件で、音声遅延を認知することができなかった可能性について述べた。これが正しいければ、音声遅延 400msec, 600msec 条件では、被験者が音声遅延に気付かずに話者交替を行おうとし、音声遅延の影響を受け、発話衝突確率が増加した可能性が考えられる。反対に、音声遅延量が 800msec, 1000msec の条件では被験者が音声遅延に気づいていたため、これを考慮に入れて話者交替を行い、結果的に発話衝突確率が下がった可能性が考えられる。そこで、被験者の発話の仕方が具体的にどのように変化したかを調査した。1 つ前の発話音声の再生が終了した後、次の発話が開始されるまでの時間を発話開始間隔とし、その平均値を音声遅延量の条件ごとに図 3.16 に示す。実験実施時のミスにより、1 グループ分のデータを集計できなかったため、ここでは残りの 9 グループ分のデータを集計した。音声遅延量 400msec の条件までは、1 つ前の発話が終了してから、平均して 800msec 後に発話開始していた。つまり、音声遅延量 0msec の条件と同じだけ間隔を空けて発話していた。しかし、音声遅延量 600msec, 800msec, 1000msec までの条件では、音声遅延量が高いほど発話開始間隔が短くなった。二元配置分散分析（群間自由度 5，群内自由度 8）を行った結果、F 値 4.61，P 値は 0.002 であり、有意水準 1 % で音声遅延量により発話衝突確率に差があるという結果が得られた。Tukey の方法により多重比較を行ったところ、有意水準 5 % のとき、音声遅延量 0msec と 1000msec, 200msec と 1000msec, 400msec と 1000msec, 600msec と 1000msec のときの発話衝突確率の平均差はそれぞれ 1.04, 0.81, 0.98, 0.91 となり、これらの条件間で発話衝突確率に有意な差が認められた。このことから被験者は音声遅延に気づき始め、音声遅延を見越して発話開始を早めることで発話衝突を回避していた可能性が考えられる。

実際に発話音声がそれぞれの参加者のスピーカから再生されるのは、図 3.16 の値に条件ごとの音声遅延量が経過した後である。これを加味して音声遅延量を足し、1 つ前の発

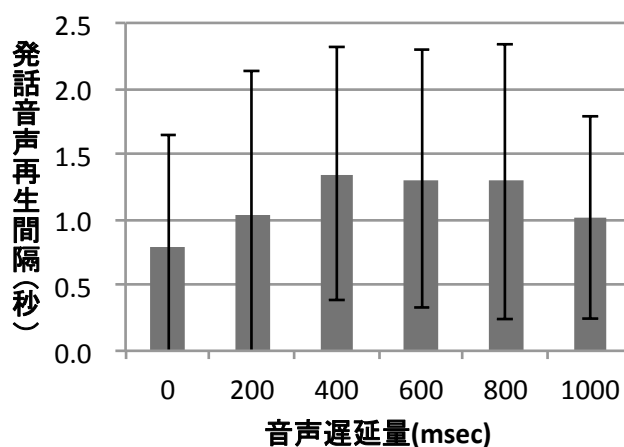


図 3.16: 発話開始間隔の比較

話が終了してから次の発話音声再生されるまでの時間を図 3.17 にまとめる．二元配置分散分析（群間自由度 5，群内自由度 8）を行った結果，音声遅延量が増えることによる発話音声再生間隔の差異は認められなかった ($P=0.19$)．これは，図 3.4 において音声遅延量が増えることによる発話回数の差が見られなかったこととも矛盾しない結果である．本実験を行った環境において，音声遅延量 0msec の条件では，被験者は一つ前の話者の発話が終わってから平均 800msec の間隔を空けて発話を開始していた．これだけの間隔があれば他の参加者が発話するか否か観察してから次の話者が発話を開始できるが，音声遅延により発話音声の再生時刻が遅れると，他者が発話するつもりがなく，自身が発話可能であると判断する可能性が高くなる．そこで発話開始を早めることで，この可能性を低くしていたことが考えられる．つまり，音声遅延量 600msec を超えたところから，徐々に参加者が音声遅延に気づき始め，その遅延量を見越して発話開始のタイミングを早めることで，発話音声再生のタイミングが，音声遅延無しの人に近づくよう努力し，結果的に発話衝突が起こりにくくなっていたと推測できる．図 3.2 の例にあてはめると，参加者 A が発話開始を早めることで，参加者 B が早めにその発話に気づき，自身の発話をやめる判断をすることができる．

3.5 快適な Web 会議システム構築へ向けた課題

本実験で用いた創出されるアイデア数に関する評価結果から，音声遅延量によって会話の効率性に差が生まれないことが分かった．しかし参加者の受ける精神的なストレスは音声遅延量によって異なることが明らかになった．精神的なストレスは長時間の，あるいは反復したシステムの利用への障壁となる．そのため，これを減少するために音声遅延を抑える，または音声遅延のある場合には発話衝突を低減する必要がある．

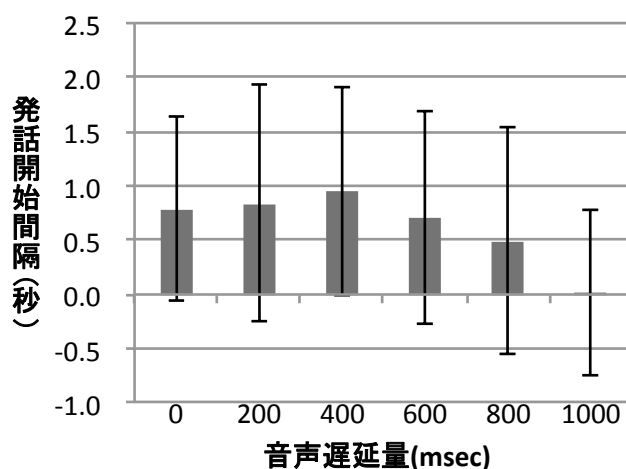


図 3.17: 発話音声再生開始間隔の比較

本実験条件の範囲では、音声遅延量が 200msec であれば、発話衝突確率、精神的ストレスは音声遅延なし条件と同等であった。音声遅延量 600msec の条件では発話衝突確率が遅延なし条件と比較して高くなった。統計的有意差こそ出ていないが、精神的ストレスを測る主観評価結果もわずかに高まっていることから、この条件でシステムを長期間使用していると精神的なストレスが蓄積してしまう可能性も考えられる。音声遅延量が 800msec、1000msec の条件では、精神的ストレスが高まり、発話衝突からの復旧が困難な可能性も示唆された。今回の実験結果とその考察から導き出される、発話衝突や、それに起因して起こる精神的ストレスを低減するための遠隔会議の要件をまとめると以下ようになる。

- (1) 音声遅延量を 200msec 以内に抑える：音声遅延量が 200msec 以内の条件では、発話衝突確率と、精神的ストレスが、音声遅延無し条件と同等に抑えられることが分かった。
- (2) (1) が実現できない場合、以下の解決方法を考える。
 - (a) 音声遅延量の影響を受けなくて済む程充分に早く到達する何らかのシグナルを伝達することで発話衝突を未然に防ぐ。具体的には、次に発話する可能性の高い参加者を予測し、あらかじめ他の参加者へ伝えることなどが考えられる。
 - (b) 音声遅延量をモニタリングできるようにする：音声遅延量が 800msec、1000msec の条件では、被験者が発話開始を早めることで、発話衝突確率が低減していた可能性について述べた。意図的に発話音声のエコーを発生させるか、音声遅延量を視覚的に提示する方法も、参加者に音声遅延が存在することを認知させることで、参加者が発話開始を早めるなどして発話衝突を防ぐ助けとなるかもしれない。

- (c) 発話権のファシリテートをする：何らかの方法により，1つ前の発話が終了する時刻を予測し，その時刻よりも音声遅延量だけ早く，システムが次の話者を指定する，もしくは参加者が発話権獲得の予約をする仕組みを作る．
- (d) 発話衝突後，その状態から復旧させるための支援をする：発話衝突が生じた際，誰が発話すべきかをシステムが指定する．

音声遅延量を 200msec 以下に抑えることが困難な状況は十分に考えられる．そのため，本論文では上記解決方法のうち，話者交替において時間的に早い段階へアプローチする (2)-(a) の方法に関して，実現方法を検討していく．

第4章 発話衝突低減手法を確立するための のアプローチ

前章では、音声遅延が高まる程、発話衝突確率が増加することを明らかにした。そして発話衝突を低減する解決方法の1つとして、音声遅延を次に発話する可能性の高い参加者を予測し、あらかじめ他の参加者へ伝えることを挙げた。そこで、対面コミュニケーションにおいて人が話者交替を円滑に行うために用いている予備動作に着目する。

まず始めに予備動作について定義する。Vargasはその予備動作として、相手に分かるように強く頷くことや、注意を引くように相槌を打つこと、組んだ足を下ろす、腕組みをほどく、身体を話者の方へ向ける、前へ乗り出す、手を挙げる動作などを挙げた[67]。これらは発話権を獲得したい参加者が意識的に行う動作であると考えられる。著者はこのような意識的に行う動作の他にも、無意識的に頷きの頻度が増えることや、相槌を打つこと、身体を前へ乗り出すこと、手を口元へ動かす動作なども、発話前に行う特徴的な非言語情報であることを明らかにした[62]。意識的か無意識的に関わらず、発話前に表出する非言語情報を本論文では予備動作と定義する。

図4.1は予備動作の分類を示す。遠隔コミュニケーションでは対面コミュニケーションと比較して、全体的に予備動作の使用頻度が低い[62]。ここで、意識的に発話権を獲得しようとして行う予備動作を「意識的な予備動作」と呼び、意識していないが発話する前に自然に現れる予備動作を「無意識的な予備動作」と呼ぶこととする。無意識的な予備動作の種類は、遠隔コミュニケーションと対面コミュニケーションにおいて大きな差異はない。これに対して、意識的な予備動作は両者の間に多少の差異が認められる。対面コミュニケーションにおける意識的な予備動作は上記のVargasによって明らかにされた予備動作があり、遠隔コミュニケーションにおいては、映像に大きな変化をもたらす動作が用いられる。具体的には、目立つように手を挙げる、手を振る、稀ではあるがカメラを手で塞ぐ動作などがある。遠隔コミュニケーションにおいて、人は映像の大きさや配置など、その時の環境に応じてより適した意識的な予備動作を学習していく。

4.1 分析

続いて、発話衝突を低減する手法を見いだすために着目している予備動作が、Web会議においてどの程度有効に用いられているかを調べた。Web会議を実施した様子をビデオで撮影し、発話衝突が起こる場面を分析した。

4.1.1 Web会議における発話衝突の例

商用Web会議システムMeetingPlaza[39]を用い、Web会議システムの使用経験がない4人の被験者に10分間の会議を実施させた。会議の議題は、4人が参加するキャンプで実施するレクリエーションを考えることとし、司会者を設けずに会議を行わせた。このうち発話頻度に偏りの少ない5分間を分析の対象とした。会議のビデオを解析し、発話と予備動作、発話の衝突を記録した。なお、明らかに相手の発話を聞いてから、それを遮るように発話することで生じる意図的な衝突は、今回の分析から省いた。この会議において観察された予備動作は以下の4つであった。

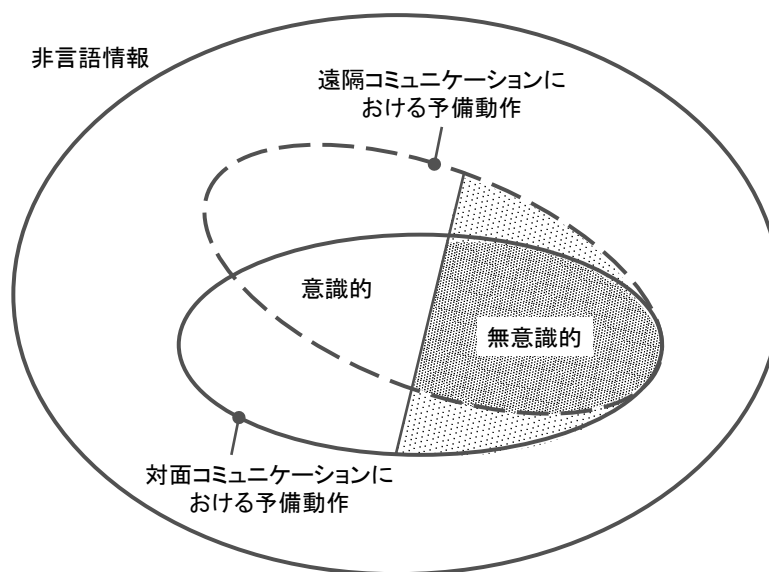


図 4.1: 予備動作の分類

- 手：口元や顔周辺へ持っていく動作
- 頭：横へ動かす動作
- 頷く
- 音声：相槌や笑いなど、発話に対するポジティブなフィードバックを返すもので、その後に発話することが多く観察された。

これらの中には意識的な予備動作、無意識的な予備動作の両方が混在していたと考えられる。上記5分間に45回の話者交替が起こり、11回の発話衝突が観測された。発話の衝突は、次の2つの状況に分類できた。

- (1) 複数の参加者が同時に発話して衝突
- (2) 発話の切れ目（呼気段落）に割り込もうとして衝突

それぞれ実際に発話の衝突が起こったシーンの具体例を挙げて解説する。

- (1) 複数人の参加者が同時に発話して衝突

1つ目の発話衝突の例を図4.2に示す。参加者1が全体に問いを投げかける発話をして終了した。参加者2～4は発話の衝突が起こる10秒前より各々、予備動作である可能性のある動作をしていた。なお、動作をしている映像が再生されているタイミングをタイムライン上に示している。参加者2はさらに3秒後にもう一度頭を動かす予備動作を行っていた。しかし後のヒアリングによると、この時参加者2～4は参

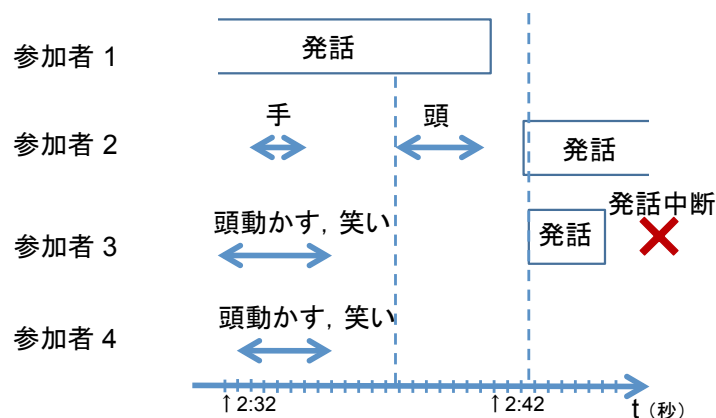


図 4.2: 複数の参加者が同時に発話して衝突する様子

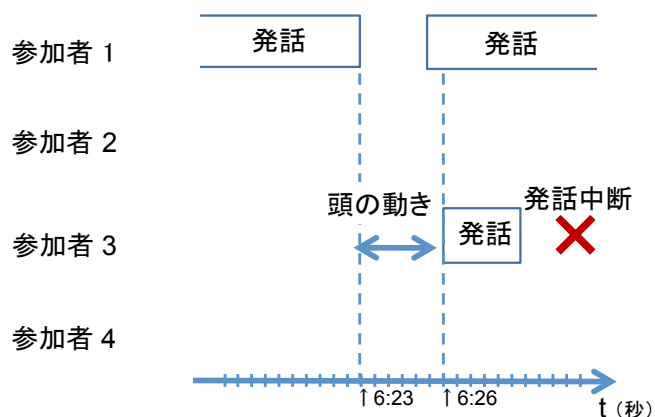


図 4.3: 発話の切れ目に割り込もうとして衝突する様子

参加者 1 の映像を見ていて、互いの動作は認知できていなかった。結果的に参加者 2 と 3 が同時に発話開始し、衝突した。参加者 3 は発話を中断し、参加者 2 は発話を継続した。

(2) 発話の切れ目に割り込もうとして衝突

2 つ目の発話衝突の例を図 4.3 に示す。まず参加者 1 が発話をしていた。参加者 3 は、参加者 1 の発話音声が届切れた瞬間を狙って発話を開始した。発話の前には頭を横に動かす予備動作をしていたが、後のヒアリングによると、参加者 1 はそれに気付かず発話を継続した。結果的に発話は衝突し、参加者 3 は発話を中断、参加者 1 は発話を継続した。

4.1.2 Web 会議における予備動作の使われ方の分析

4.1.2.1 手順

前節の発話衝突場面の例では、いずれも予備動作を行ってから発話したにもかかわらず発話の衝突が起こっていた。これらの予備動作の活用度合いが Web 会議と対面会議でどの程度異なるかを確かめるため、前節で説明したものと同様の会議を、対面会議で実施し比較した。対面会議においても発話、予備動作、発話の衝突を記録した。

ある参加者が予備動作を行ってから発話するまでの間に、別の参加者の発話が挿入されることは十分に考えられる。しかし現状、予備動作をした時の発話意図が、いつまで持続したかを測る手段はない。そのため今回の分析では、予備動作である可能性のある動作を行った参加者が、すぐ次のターンに発話権を獲発話を開始したかどうかに着目し、集計を行った。

4.1.2.2 予備動作の出現回数と効果

対面会議では5分間で85回の話者交替が起こり、1回の発話衝突が観測された。表4.1は、対面会議とWeb会議において予備動作である可能性のある動作が観察された回数と、その動作を行った後に発話した回数、そして後者を前者で割った値を動作後の発話確率として比較し、まとめたものである。Web会議では、この予備動作である可能性のある動作の合計回数は78回であり、対面会議での195回と比較して半分以下であった。

この、予備動作である可能性のある動作の回数には、発話意図を伴う予備動作の回数と、発話意図を伴わない、その他の動作の回数が含まれる。現状、予備動作とその他の動作を見分けることは容易ではない。しかし、ある参加者が発話をした場合、その直前に行っていた動作は予備動作である。したがって、予備動作である可能性のある動作をした後に発し、かつそれが衝突しなかった確率を求めることで、予備動作後の発話の非衝突確率を求めることができる。図4.4に予備動作後の発話の非衝突確率を示す。対面会議では予備動作後の発話はほぼ100%に近い確率で成功しているが、Web会議ではどれも80%を切っていた。表4.1から、予備動作全体に占める音声の割合は対面会議で32%(44回/195回)であるのに対し、Web会議では23%(25回/78回)と多かったことが分かる。しかし、その予備動作の可能性のある動作後でも20%の発話が衝突している。手や頭の動きによる予備動作後の発話の非衝突確率は40%、60%と低く、このことから、Web会議では映像チャンネルを通じて伝達される予備動作は認識されづらいことが推測できる。

以上をまとめると、今回の分析範囲内では、Web会議では対面会議と比較して予備動作である可能性のある動作の回数が半数以下に減少し、予備動作後の発話の非衝突確率も大きく減少する傾向があった。

4.1.2.3 予備動作出現の特徴

今回の分析範囲内では予備動作表出の特徴として以下の傾向が見られた。

表 4.1: 動作後の発話確率

予備動作である可能性のある動作	対面会話での発話確率 (発話回数 / 動作回数)	Web 会議での発話確率 (発話回数 / 動作回数)
手	61 % (28/46)	73 % (8/11)
頭	55 % (41/75)	43 % (15/35)
頷き	57 % (17/30)	71 % (5/7)
音声	57 % (25/44)	60 % (15/25)
合計	57 % (111/195)	55 % (43/78)

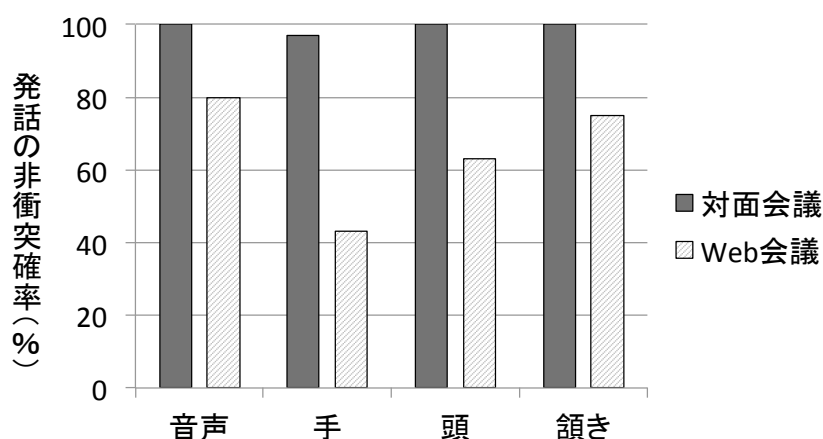


図 4.4: 予備動作後の発話の非衝突確率

(1) 予備動作の種類により，その後の発話確率が異なった（表 4.1）

(2) 最も多くの予備動作をした参加者が発話する割合が高った

図 4.5 に，Web 会議での 45 回の話者交替のうち，直前の発話中に行った予備動作多さの順位と発話の関係を示す．今回の分析では，半分以上の話者交替時に，最も多くの予備動作を行った参加者が発話していた．なお，最も多く予備動作を行った参加者が発話せず，それ以外の参加者が発話した 7 回のうち 2 回は，呼びかけに反応するなど，明らかに次話者を指定された場面であった．

(3) 予備動作の種類によって，出現してから発話開始するまでの間隔が異なった

手と頭による予備動作は，表出後すぐに発話することが多く観察された．音声による予備動作は，動作後そのまま発話に移ることが多かった．それらに対して頷きは，直前の発話の中ごろから表出することが多く観察された．

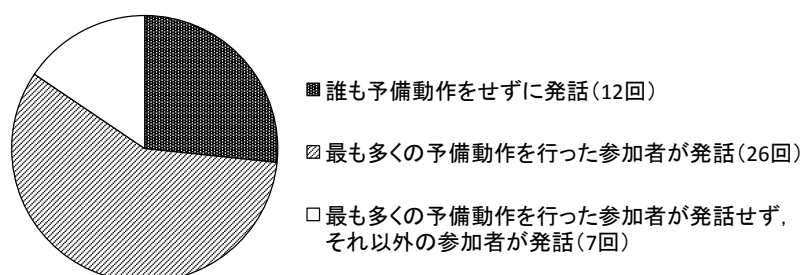


図 4.5: Web 会議における予備動作回数と発話の関係

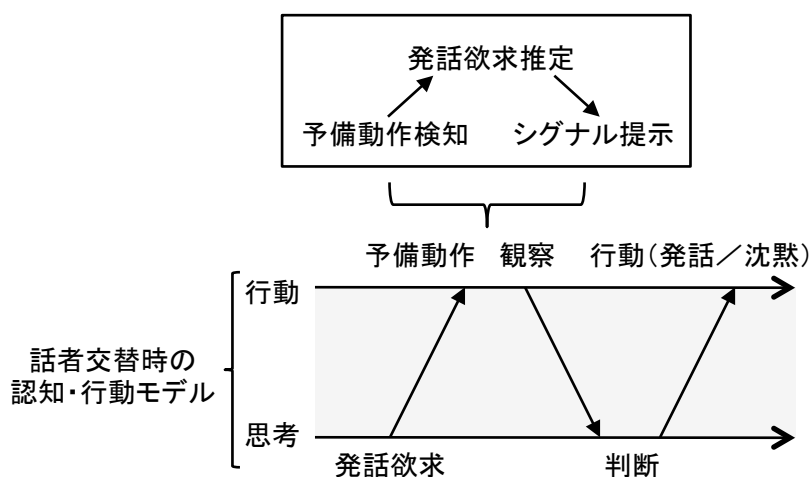


図 4.6: 話者交替時の認知・行動モデルにおいて本提案コンセプトが支援するステップ

4.2 発話衝突を低減するための基本コンセプト

本研究の目的は、Web 会議での発話衝突を低減することにより、自席のデスクトップパソコンやノートパソコンからでも、創造会議を快適に行える遠隔会議環境を実現することである。図 3.1 の話者交替時の認知・行動モデルは、対面コミュニケーションにおいて滞りなく流れることが考えられる。3 章では、このモデルが音声遅延によって影響を受け、発話衝突確率が高まるという仮説を検証した。そして前節まででは、発話開始よりも充分前もって表出する予備動作が、Web 会議においてうまく相手に伝わりづらいことが分かった。

そこで、図 3.1 の予備動作ステップにおいて表出したある参加者の予備動作を、他の参加者が、観察ステップにおいて認知しやすくする方法を提案する。図 4.6 のように、予備

動作ステップから観察ステップへの情報の流れを補助するような、予備動作検知ステップ、発話欲求推定ステップ、シグナル提示ステップを挿入する。具体的には、予備動作検知ステップで会議参加者の予備動作を検知し、その情報を基に参加者の発話欲求を発話欲求推定ステップにて推定する。そして他の参加者へとシグナルを提示することで、予備動作および発話欲求の伝達を支援する。シグナルとは、参加者への注目を集めるための何らか（映像や音声など）の刺激を指す。

前章で述べたように、インターネットを介して行う Web 会議は遅延の影響を大きく受ける。そのため上記シグナルが他の参加者のもとへ到達するまでも音声と同程度の遅延が生じる。しかし、4.1.1 節で述べたように予備動作は発話開始よりも充分早く表出することが多い。そのため遅延を加味しても、このシグナルは発話開始以前に他の会議参加者へと提示され、会議参加者各々が発話するか否かを判断する助けとなり、発話衝突を回避することにつながると考えられる。

第5章 次話者候補提示法

5.1 次話者候補提示手法の提案

前章での分析の結果、普段我々が対面コミュニケーションで話者交替を円滑にするために活用している予備動作が、Web 会議では認知されづらいことが分かった。そこで著者は前章で参加者の予備動作をシステムが検知し、その情報を基に発話欲求を推定し、他の参加者へと伝達するコンセプトを提案した。そこで本コンセプトを実現するために、次話者候補提示手法を提案する。まず参加者の映像と音声から予備動作を検知する。ここで検知する予備動作は、意識的なものと無意識的なもの両方が対象である。

発話欲求推定ステップでは発話欲求の度合いを推定するが、現時点ではそれを確実に推定することは困難である。そこで本手法では、提案するコンセプトを実現するシンプルな手法として、すべての参加者の中で、最も推定結果の高い参加者を次の話者候補（次話者候補）として選定する。随時更新される発話欲求度合いから次話者候補を1人選択し、それをすべての参加者へ伝えるために、次話者候補の映像を強調して表示する。時話者候補に選定された参加者は他の参加者からの注意を集め、自然に発話を開始することができる。ここで複数の参加者が同意に発話したい欲求を抱いた場合でも、次話者候補に選定されなかった参加者が次話者候補へ発話権を譲ることで、発話衝突確率を低減させることができる。

5.2 オズの魔法使い実験による検証

前節で述べた提案手法により、Web 会議における発話の衝突確率を低減できる見込みがあるかどうか、「オズの魔法使い実験 [25]」により確認した。オズの魔法使い実験とは、システムの未完成部分を人手で補って動かし、そのシステムの有効性を検証する実験方法である。本実験では提案手法のうち、参加者の予備動作を検知し、次話者候補を選定する部分を、人の手によって行う、オズの魔法使いシステムを構築した。魔法使い役（実験者）は、Web 会議映像・音声をリアルタイムで視聴し、予備動作を観察する。そしてそれを基に次話者候補を選定する。本実験では予め参加者ごとに ID を振っておき、その ID に対応するキーをタイプすると、相当する参加者映像に青い枠が表示されるようにした。つまり、Web 会議を実施する際に、参加者の見ている映像の中で、魔法使い役が選定した次話者候補の映像に青い枠が表示される。

上記オズの魔法使い条件と比較するために、システムにより5秒に1度、ランダムに次話者候補を選定し、その参加者映像に青い枠を表示させるシステムを構築した。以上2条件に、従来通りの Web 会議システムを用いた、枠表示無し条件を加えた3条件にて比較実験した。4人の参加者で、それぞれの条件につき、5分間の会議を実施した。議題は「テレビの新機能を考える」とした。実験を行った順序は「枠表示無し」「枠表示有り（魔法使い）」「枠表示有り（ランダム）」とし、後2つの条件ではどちらとも「次に発話しそうな参加者に枠が表示される」と教示した。実験結果は図 5.1 の通り、枠表示有り（魔法使い）条件で発話衝突確率を下げる事ができた。ランダムに枠を表示した条件では発話衝

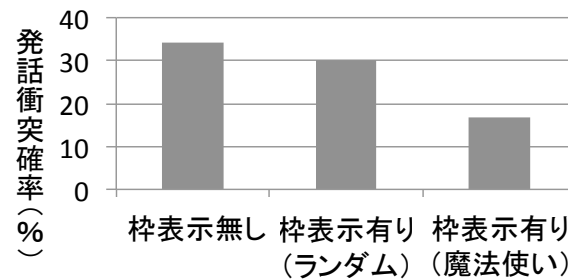


図 5.1: オズの魔法使い実験の結果

突確率を減らすことができなかった。この条件では、発話したい欲求のない参加者に枠が表示される割合が多く、被験者がそれを無視して会話をした可能性が考えられる。

5.3 プロトタイプ実装

提案概念を実現するプロトタイプシステムを実装した。3章での分析を基に、予備動作を検出し、「次話者候補」をすべての参加者に示す機能を実装し、MeetingPlaza に組み込んだ。

5.3.1 予備動作の検知

Web 会議システムで使用する Web カメラ、ヘッドセットから入力される映像、音声に基づいて以下の予備動作をリアルタイムに検知する機能を実装した。

- 手：手を上下に動かす動き
- 頭：頭が上下左右へ移動する動き
- 頷き：顔を縦に動かす動作
- 音声：話者でないときの発声

実装では Web カメラの正面に参加者が 1 人座り、背景が変化しない環境を想定した。顔認識には OpenCV に組み込まれた認識機能を使用した。なお、プロトタイプシステムにおける参加者映像の解像度は 320x256、リフレッシュレートは 30fps であった。各動作の精密な音声・画像処理技術自体は本研究の対象外である。

- 手
映像を縦に 4 分割したうち、左右両端いずれかの領域で、フレーム間差分をとり、変化が生じた場合、手の動きと判定させた。

- 頭
顔領域の中心点の x, y 座標のうちいずれかが、3 フレームの間に 30 ピクセル以上移動した場合、頭の動きと判定させた。
- 頷き
顔領域の中心点がフレーム間で x 座標 ± 2 ピクセル以内、かつ y 座標 ± 30 ピクセル以内で移動した場合、頷きと判定させた。
- 音声
最大 1 人の参加者が話者として指定される。いずれかの参加者のマイクから音声が入力されている場合、最も先に音声入力のあった参加者が話者として指定される。話者の音声入力がなくなった場合、次に最も先に音声入力のあった参加者が話者として指定される。話者でないときの音声入力を、すべて音声による予備動作と判定させた。

5.3.2 次話者候補の選定方法

普段我々が対面コミュニケーションをする際には、予備動作を認知し、現発話とのタイミング、いくつかの動作の組み合わせ、などを複合的に処理して次話者を判断し、発話権授受を行っていると考えられている [11]。普段我々が対面コミュニケーションをする際には、予備動作を認知し [67]、現在の発話の切れ目を予測し、誰が次の話し手になるのかという次話者の選択も行っている [11] と考えられる。しかし今回はまず 3 章で行った分析により得た下記項目を要素として満たす、次話者候補選定アルゴリズムにより、提案手法の有効性を評価する。

- 予備動作の種類によって、その後の発話確率が異なる
- 複数の予備動作をした参加者は次話者になる確率が高まる
- 予備動作の種類によって、その後の発話の「予備」動作として有効な時間は異なる

予備動作毎に、その動作後の発話確率に応じたスコアと持続時間を付与（表 5.1 参照）した。発話したいという意図（発話意図）があったとしても、多人数会話であれば必ずしも発話権を獲得できるわけではない。そのため、今回の分析で得られた予備動作後の発話確率よりも、発話意図は高いことが予想される。しかし、今回の実装においては、予備動作後の発話確率が高いほど、その動作をしたときの発話意図が高いものと仮定してスコアを定めた。

持続時間に関して、手と頭による予備動作は、出現後すぐに発話することが多く観察されたため、1 秒間とした。頷きは、現発話の中ごろに出現してから発話することが多く観察された。分析対象とした会議では 6 秒間程度の発話が多く観察されたため、頷きの持続時間は半分の 3 秒とした。音声による予備動作は、その予備動作後そのまま発話に移ることが多かった。そのため持続時間は音声入力のある間とした。

表 5.1: 発話の予備動作のスコア付け

予備動作	予備動作後の発話確率	スコア	持続時間
手	73 %	8	1 秒
頭	43 %	4	1 秒
頷き	71 %	7	3 秒
音声	60 %	6	音声入力のある間

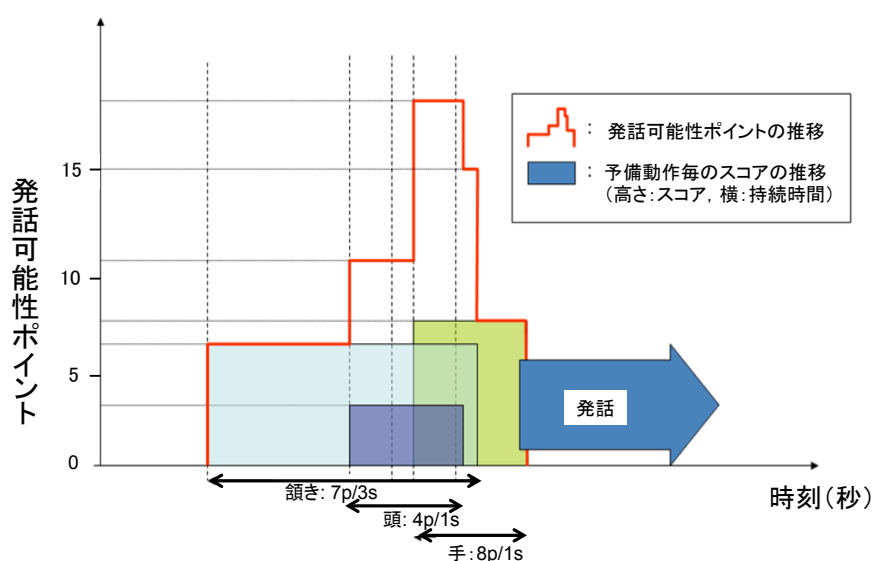


図 5.2: 予備動作毎のスコアの推移と発話可能性ポイント

ここで定める時間やスコアは最初のプロトタイプのための初期値となるもので、今後実験や検証を繰り返すことでより妥当な値を求めていく必要がある。

ある時刻における発話可能性ポイントは、持続時間中にあるスコアの線形和である。すなわち、ある参加者の時刻 t における発話可能性ポイント $P(t)$ は、スコア $s(i)$ と持続時間関数 $d(t, i)$ により、次式のように算出される。

$$P(t) = \sum_{i=1}^n (s(i) \times d(t, i)) \quad (5.1)$$

ここで、 i は予備動作を表す識別子で、表 5.1 によれば、例えば 1: 手, 2: 頭, 3: 頷き, 4: 音声である。持続時間関数 $d(t, i)$ は、予備動作 i が検知された時刻 t_0 から時刻 t が持続時間内であれば 1 を、時間外であれば 0 を取る関数である。

図 5.2 は参加者の行った予備動作に応じて発話可能性ポイントを求める例を示している。発話可能性ポイントは、実際にはより複雑な式によって求まると推測されるが、本実装で

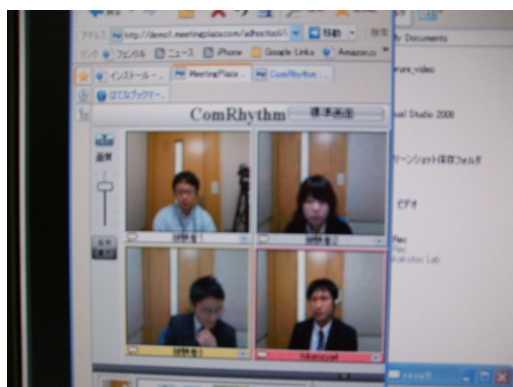


図 5.3: プロトタイプシステム実行画面

は上記のような計算方法を採用した．

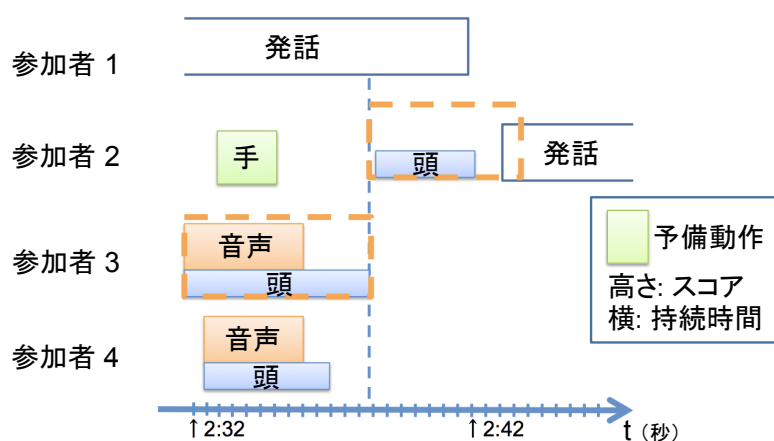


図 5.4: 発話可能性ポイントの推移と次話者候補選択の例

ある時刻での発話可能性ポイントが最も高い参加者を次話者候補として選択した．上記基準により選択された次話者候補の参加者映像に黄色い枠（以降，次話者候補提示枠と表記）をつけることで，全参加者へと次話者候補であることを示した．MeetingPlaza には元々の機能として，話者に赤色の枠がつく仕様となっている．黄色から赤への遷移が我々にとって信号機でなじみ深く，認知しやすいと考え，話者の前段階である次話者候補を表す色にも黄色を採用した（図 5.3）．動作の例として，4.1.1 節で挙げた発話の衝突場面（図 4.2）に，本手法を適用していた場合の動作を説明する（図 5.4）．参加者 1 が発話している間，他の参加者 2～4 にはそれぞれ行った予備動作に応じて，発話可能性ポイントが計算されている．ここで示されているタイムラインの前半部では参加者 3 の予備動作が多く，最も発話可能性ポイントが高いため，次話者候補を示すために，参加者 3 の映像に黄色の枠が表示されている．ところがタイムラインでの中盤になると，参加者 2 が頭を

動かしたことによってそのスコアが足しあわされ、この時点で最も発話可能性ポイントが高くなった。すると今度は参加者2の映像に黄色の枠が表示される。そして参加者1の発話が終了した際に、参加者3と4は、枠に色が付いていた参加者2に発話権を譲り、参加者2はそのまま発話を成功させることができる。著者らが MeetingPlaza を用いて往復の音声遅延を測定したところ 600msec であったが、前章の分析より、予備動作はこの遅延を加味しても充分前もって出現するため、この手法は発話の衝突を回避するために有効に機能すると考える。

5.4 実験

5.4.1 目的

プロトタイプシステムを用いて、本提案コンセプトの有効性を検証する。具体的には、次話者候補の参加者映像に次話者候補提示枠を表示することで発話の衝突確率が低減できるかを確かめる。

5.4.2 手順

プロトタイプシステム（次話者候補を示す次話者候補提示枠有り）と、従来の Web 会議システム（次話者候補提示枠無し）との比較実験を行った。Web 会議システムには MeetingPlaza を使用した。それぞれのシステムを用いて被験者に創造会議を実施させ、発話の衝突確率を比較した。20代～30代の男女30人の一般被験者を募り、3人ずつの10グループを構成し実験を実施した。被験者のグループメンバー3人は互いに親しい知り合いであることを条件とした。会議の議題は「テレビの新機能を考える」とした。グループで話し合ってアイデアを出していき、グループ内に1人定めた書記係に、そのアイデアを記録させた。提案手法である次話者候補提示枠有り条件と、従来手法である次話者候補提示枠無し条件の2つを交互に行い、その順序はランダムとした。被験者10グループをさらに5グループずつの2群に分け、以下の通り、群ごとに次話者候補提示枠有り条件において教示する情報を変えた。片方の5グループでは、「次話者候補提示枠は次に話しそうな参加者映像に表示される」とだけ教示し、もう片方の5グループには、上記内容に加えてシステムが検知している具体的な予備動作まで教示した。前5グループは「動作非教示群」と記述し、後5グループは「動作教示群」と記述する。

1つのグループあたり合計14分間の会議を実施し、前後半7分に分け、次話者候補提示枠有り、無しの条件をそれぞれ実施した。条件の順序はランダムに設定した。14分間を通して議題は継続させた。いずれの条件においても、映像のリフレッシュレートは30fpsに設定し、ある参加者が発声してから他の参加者へ伝わるまでの往復の音声遅延量は400ms程度であった。実験中の会議システムの映像と音声をビデオカメラで記録し、解析を行った。

教示する内容は統一することが望ましかったが、どちらか一方を合理的に選択することができなかった。動作非教示群では、検知される動作を知らないユーザが自然にシステム

を使用し、次話者候補提示枠に誘導され、話者交替を行うという状況である。自身の振る舞いを意識せず、自然に会議をするため、発話回数は次話者候補提示枠無し条件と同等となり、発話衝突確率が減少することを予想した。一方動作教示群では、このシステムを商用化する際に説明書に記載される程度の内容、すなわち頷き、手、頭、音声といった具体的な予備動作の種類を教示してある。このため、積極的に予備動作を使用するようになり、次話者候補提示枠無し条件と比較して、発話衝突確率が減少するが、意図的な動作を挟む機会が増え、発話回数は減少することを予想した。

5.4.3 結果

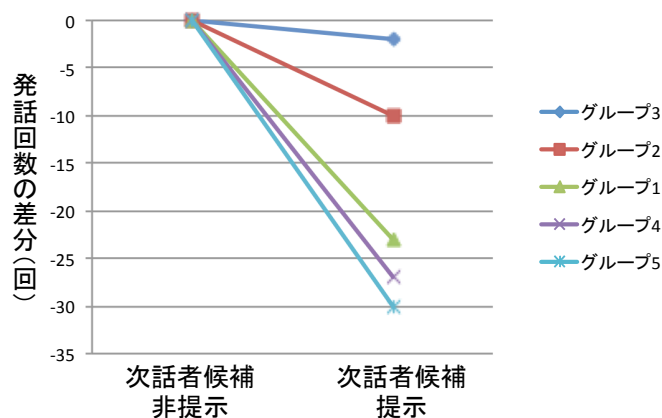


図 5.5: 動作非教示群における、次話者候補提示有無による発話回数の比較

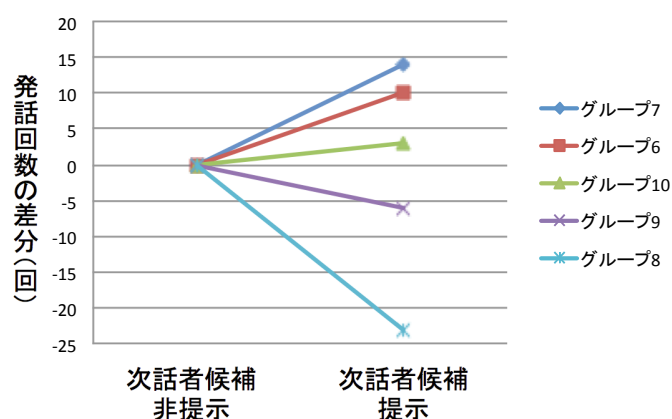


図 5.6: 動作教示群における、次話者候補提示有無による発話回数の比較

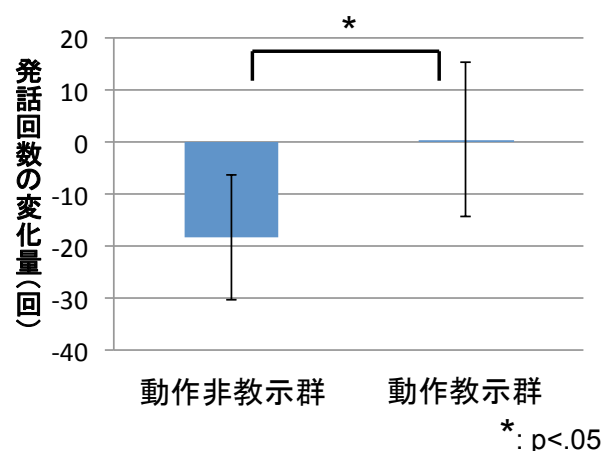


図 5.7: 次話者候補提示有無による発話回数増加量の比較

それぞれの条件ごとに7分間の発話回数，発話衝突回数を記録した．また，発話衝突回数を，発話回数と発話衝突回数の和で割った値を，発話衝突確率として求めた．

発話回数について，次話者候補提示枠無し条件（従来手法）を基準としたときの，次話者候補提示枠有り条件（提案手法）での増減を，動作教示群に関しては図 5.5 に，動作教示群に関しては図 5.6 にまとめた．動作非教示群では，5つのグループすべてが，次話者候補提示枠有り条件において，発話回数が減少していた．一方，動作教示グループでは，次話者候補提示枠無し，有りの条件による増加，減少の傾向は見られなかった．次話者候補提示枠無し，有り条件間の発話回数の増減の平均値を，動作教示・非教示ごとに算出し図 5.7 にまとめる．平均値の比較をt検定にて行ったところ，有意差5%で，動作非教示群における，次話者候補提示枠有無での発話回数の減少は，動作教示群のそれよりも大きくなることが分かった．

発話衝突確率についても同様に，次話者候補提示枠無し条件を基準としたときの，次話者候補提示枠有り条件での増減を，動作非教示群に関しては図 5.8 に，動作教示群に関しては図 5.9 にまとめた．動作教示群では，次話者候補提示枠無し，有りの条件による増加，減少の傾向は見られなかった．動作教示群では，5つのグループすべてが，次話者候補提示枠表示条件にて，発話衝突確率が減少していた．動作教示群と，動作非教示群の間で，発話衝突確率の次話者候補提示枠無し，有り条件間の差の平均値を算出し，それに-1を乗じたもの値を図 5.10 にまとめる．これはすなわち，動作非教示群，動作教示群のうちどちらが，次話者候補提示枠を表示したときに，より発話衝突確率を低減できたかを比較した図である．両平均値の比較をt検定にて実施したところ，有意差5%で，動作教示群の方が，次話者候補提示枠を表示したときに話衝突確率が減少することが分かった．

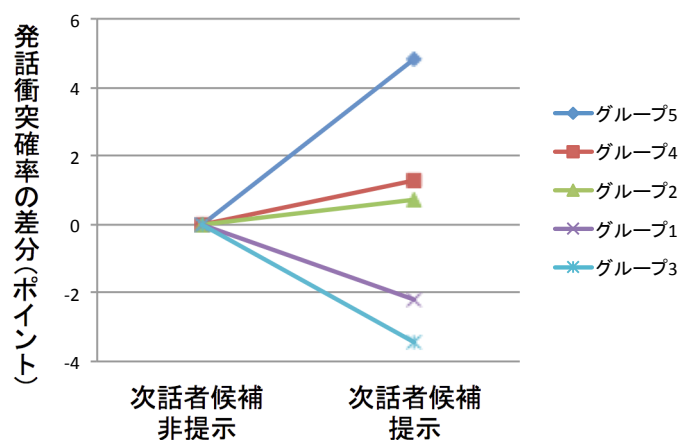


図 5.8: 動作非教示群における，次話者候補提示枠有無による発話衝突確率の比較

5.5 考察

本実験の目的は，次話者候補を強調することにより，発話が衝突する確率を低減するという，本提案コンセプトの有効性を確かめることであった。「次に最も発話しそうな参加者に次話者候補提示枠が表示される．その判定のために，頷き，手の動き，頭の動き，相槌が検知されている」と教示した「動作教示群」においては，発話衝突確率が減少することが確かめられた．しかし，「次に最も発話しそうな参加者に次話者候補提示枠が表示される」とだけ教示された「動作非教示群」において，発話衝突確率は減少しないという結果が得られた．以下にこの原因を考察する．

図 5.11 のように，発話欲求が生じてから，発話にいたる流れを考える．具体的には，ある参加者に発話したいという欲求が生じ，その参加者が予備動作を行う．そしてその予備動作をシステムが検知する．そこから次話者候補として選定された場合，その参加者の映像に次話者候補提示枠が表示され，その参加者が判断をし，最終的に発話にいたる（もしくは，発話しない）．この流れに沿って発話するとき，発話欲求ステップで生じた発話欲求を，発話ステップへと 100 % 反映することができれば理想的である．しかし現実には，それぞれのステップ間の矢印区間になんらかのノイズが挟まれ，右のステップへ進むほど，発話欲求を反映できる確率が減少していく．発話欲求ステップと予備動作ステップの間には，参加者の性質やその時の状況が影響し，予備動作を行うか否かが異なる．予備動作ステップと予備動作検知ステップの間には，予備動作の検知精度が影響する．予備動作検知ステップとシグナルステップの間には，次話者候補の選定精度が影響する．シグナルステップと発話ステップの間には，参加者がシグナル（次話者候補提示枠）を信用し，次話者候補提示枠が表示された参加者が実際に発話するか否か（次話者候補提示枠表示後発話確率）が影響する．

今回，正確な発話欲求は測定できていない．そのため，参加者が予備動作である可能性のある動作を行った際に，それが予備動作であるか否かを測定することはできなかった．

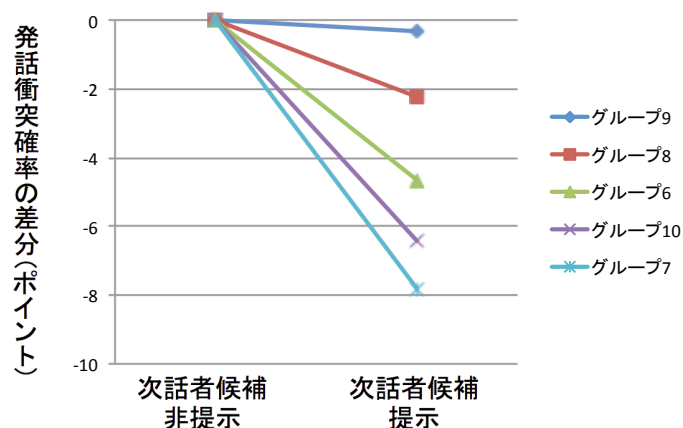


図 5.9: 動作教示群における，次話者候補提示枠有無による発話衝突確率の比較

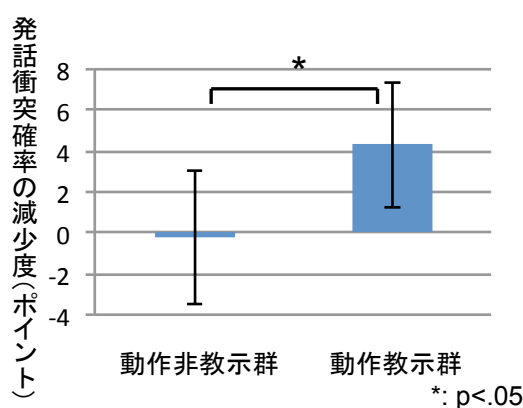


図 5.10: 時話者候補提示有無による発話衝突確率減少度の比較

したがって，本プロトタイプシステムにおける予備動作の検知精度と次話者候補の選定精度を正確に測定する手段はない．本実験とは別に，予備動作である可能性のある動作の検知精度を求めたところ，各々，手は78.3%，頭は65.8%，頷きは64.2%，音声は98.3%であった．

また，次話者候補提示枠表示後発話確率を，ビデオ解析データを基に，グループごとに算出した．次話者候補提示枠表示後発話確率と，次話者候補提示枠無し・有り条件間での発話衝突確率の減少度（発話衝突確率減少度）の関係を散布図に表した（図5.12）．この2つの値の相関を求めたところ，相関係数は0.47で，中程度の相関があることが分かった．このことから，次話者候補提示枠を使って必ずしも発話できているわけではないが，次話者候補提示枠を使って発話する割合が高いほど，発話衝突確率は減少する傾向があることが分かった．

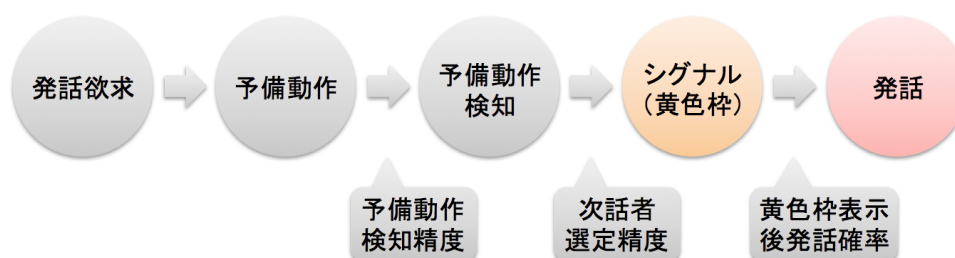


図 5.11: 発話欲求が生じてから発話するまでのステップ

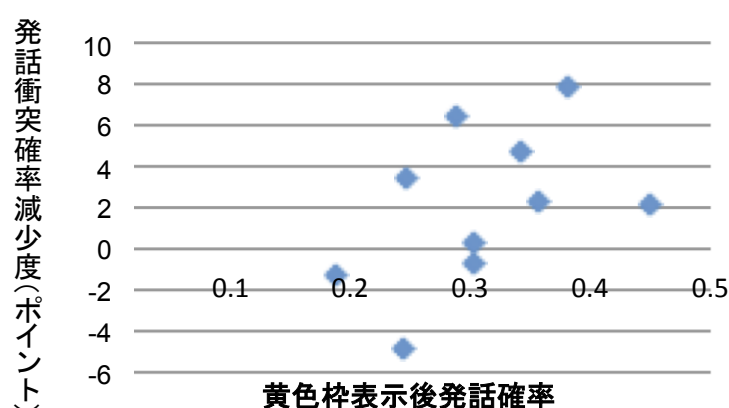


図 5.12: 次話者候補提示枠表示後発話確率と発話衝突確率減少度の関係

次話者候補提示枠表示後発話確率の、動作教示群5グループでの平均は33.4%で、動作非教示群5グループでの平均は28.6%であった。このことから、動作教示群の方が、次話者候補提示枠を使って発話する割合が高かったことが分かる。言い換えれば、動作教示群の方が、より次話者候補提示枠というシグナルを信用して発話したと考えられる。動作教示群の方が、動作非教示群よりも、次話者候補提示枠条件において発話回数が増えていることから、動作教示群では、1) 話者が、次話者候補提示枠の表示されている他の参加者へ発話権を譲ったこと、2) 次話者候補提示枠の表示された参加者が躊躇せず発話したこと、などが考えられる。動作教示群において参加者が次話者候補提示枠を信用することができた理由としては、具体的な予備動作を教示されることで、積極的にその予備動作を行い、図5.11の発話欲求ステップと予備動作ステップ間のノイズが減少し、結果的に、次話者候補提示枠というシグナルが発話欲求を反映できる確率が高まったことが推測される。

5.6 課題

本手法を用いることで、発話衝突を低減できることが明らかになった。発話衝突確率をさらに低減させるためには、予備動作検知精度を上昇させる必要がある。図 5.11 の流れにおいて、発話欲求ステップからシグナルステップまでの精度が高まる程、参加者からのシグナルに対する信用が高まる。するとそれを使用する頻度が高まり、発話衝突を低減できる可能性が高まる。十分に予備動作の検知精度を高めることで、具体的な予備動作の教示・非教示に関わらず、次話者候補提示枠を利用して発話権の授受を行い、発話衝突確率を低減できるようになり、より使い易い会議システムを実現することが期待できる。しかし、現時点では予備動作候補を正確に検知することができたとしても、そこから発話欲求を正確に把握することは困難である。この課題を解決することが、さらに円滑な話者交替を実現するためには不可欠である。

第6章 発話欲求伝達手法

6.1 発話欲求伝達手法の提案

前章において、Web カメラで取得した画像を基に、予備動作である可能性のある動作を検知し、そこから推定される、最も次に発話する可能性の高い参加者を、次話者候補として他の参加者へ強調して提示する手法を提案した。現段階で、予備動作である可能性のある動作の検知では、それが発話欲求を伴う予備動作であるか否かを正確に判断する方法は見いだせていない。予備動作候補としてこれまで著者らが挙げてきた、「頷き」「挙手」「手を顔周辺へ動かす動作」などをシステムが検知しても、前述のように、その動作をした参加者に発話欲求があるか否かを正確に把握することはできない。例えば「大丈夫」という意思を表す動作や「賛成の人は？」などの質問に対して手を挙げる動作などは、必ずしもその後に発話を伴うとは限らない。しかしこれらの動作と発話欲求のある「挙手」をシステムは正確に区別する事はできない。したがって、ある参加者の発話欲求の度合いを正確に特定する手段は見いだせていない。また、システムが次話者候補を1人に絞って提示する仕組みでは、発話欲求が正確に分からない場合、図 5.11 の判断ステップで、参加者に誤った判断をさせる可能性を多く含む。

参加者の行う予備動作候補を正確に検知しても、そこから発話欲求を正確に特定することは困難である。そこで、参加者へ推定されている自身の発話欲求度合いをリアルタイムにフィードバックする発話欲求伝達手法を提案する。参加者自身が動作を調整することで、自身の抱えている発話欲求とシステムに推定されている発話欲求の度合いをすり合わせることができ、発話欲求度合いと表示されるシグナルの整合率を高めることができる。本手法のポイントは、参加者が普段の会話で話者交替に使用している予備動作を検知して発話欲求の推定をすること、そしてシステムに推定された発話欲求の度合いと実際の発話欲求の度合いの整合性を、参加者自身が動きを調整することで近づけていく点である。すなわち、参加者はシステムに発話欲求を正しく推定されるように、意識的な予備動作を学習していくこととなる。

本手法を実現するために、以下の3つの項目を検討する。

(1) 予備動作候補を検知する精度（予備動作候補検知精度）の向上

発話欲求度合いを推定するために用いる、予備動作候補の検知精度を高める必要がある。そのために、センシングデバイスにより、参加者の頭、首、肩関節、肘関節などの3次元座標位置をセンシングし、その位置関係や動きを基に、それらの動作を検知する。センシングデバイスには、本稿では、2次元画像とその奥行き情報を高い精度で検出できる Kinect センサ [35] を採用する。

(2) 検知された予備動作候補から推定される発話欲求を、実際に参加者の抱く発話欲求へと近づける仕組みの構築

その時々検知された予備動作候補を基に、システムに推定されている発話欲求度合いを表すインジケータ（発話欲求フィードバックインジケータ）を表示する。これにより、表示されている発話欲求フィードバックインジケータが、実際の発話欲求よりも高い場合、随意的に動作を控え、発話欲求フィードバックインジケータを下

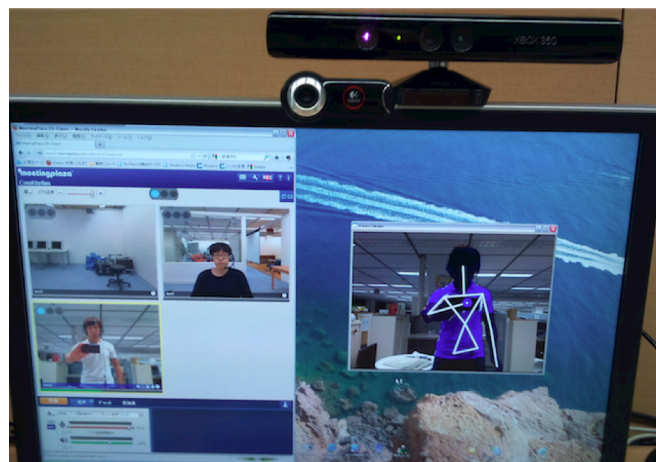


図 6.1: 発話欲求を伝達可能な遠隔会議システム実行画面

げることが可能となる．逆に，実際よりも発話欲求が低く推定されているとフィードバックされた場合には，発話欲求フィードバックインジケータを確認しながら予備動作を行うことで，推定される発話欲求度合いを引き上げることができる．人は普段の対面した会話において，発話欲求のある場合には何らかの予備動作をし，そうでない場合には動作を行わない．この予備動作はボタンを押す動作などとは異なり，普段対面会話において話者交替する際に使用しているものなので，Web 会議中の話者交替においても自然に取り入れられると考える．また発話欲求のない場合には予備動作を抑えることも参加者にとって自然であると考えた．

- (3) 参加者自身により調整された発話欲求を伝達し，他の会議参加者に対して正確な判断材料となるシグナルの提示

それぞれの参加者の発話欲求がある閾値を超えたとき，参加者間で誰に発話欲求があるか否かを確認できるようにする．すなわち，発話欲求があると判定された参加者の映像を強調し，他の参加者へ示すようにする．本提案手法では，同時に強調される参加者の数に制限を設けず，自身が発話するか否かの判断は参加者へ委ねることとした．例えばある参加者に発話欲求が生じた際に，他の参加者の映像が先に強調されていた場合，その参加者は自身の発話を控えて譲ることで発話衝突を回避することができる．

6.2 発話欲求伝達モジュールの実装

会議システム全体の系としては，図 6.2 のようになる．商用 Web 会議システムである MeetingPlaza[39] と連携できるように発話欲求伝達モジュールを実装した．各クライアントマシンにおいて，発話欲求伝達モジュールはセンサから得られた情報を基に各参加者の

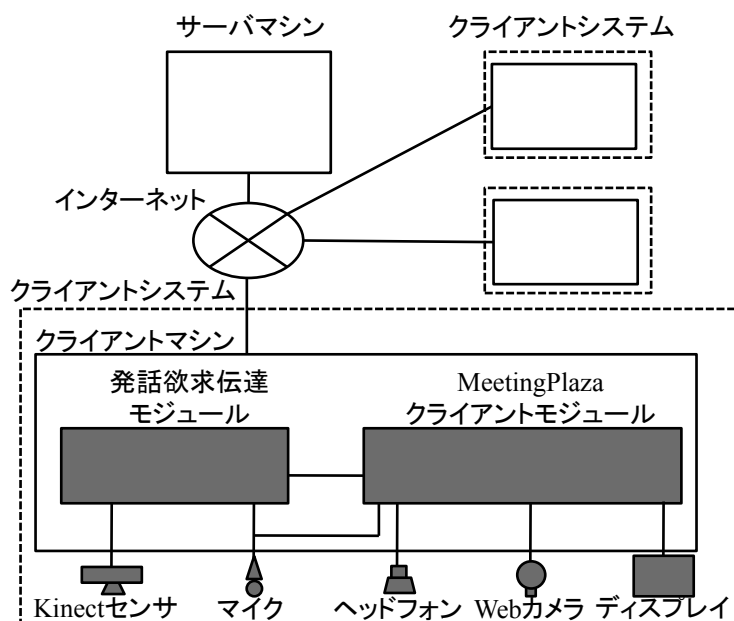


図 6.2: 会議システム全体の構成図

予備動作候補を検知し、発話欲求フィードバックインジケータや、参加者映像へ黄色い枠を表示させる要求を MeetingPlaza クライアントモジュールへ渡す。MeetingPlaza クライアントモジュールはこれらの情報をインターネット経由でサーバマシンへ送る。サーバマシンは、受信した情報をすべてのクライアントマシンへブロードキャストする。以上の流れによってクライアントマシン間で情報が共有され、発話欲求フィードバックインジケータや黄色い枠の状態が同期される。Web 会議システム画面上に表示される映像は Web カメラから取得し、マイクから取得した音声とともに、MeetingPlaza クライアントモジュールへ取り込まれ、後は前述と同様の流れで、各クライアントマシンへ共有される。

発話欲求伝達モジュールは、予備動作候補検知部、発話欲求推定部、そして発話欲求伝達部から成る。

6.2.1 予備動作候補検知部

著者は Vargas の知見 [67] を基に、「手を挙げる」、「手を顔周辺へ持っていく」、「頷く」、「身体を前後に動かす」が Web 会議中の発話前に行う予備動作であることを明らかにした [62]。このうち特に「手を挙げる」、「手を顔周辺へ持っていく」、「頷く」の3つが、参加者が意識的に行う予備動作であることが分かった。これらを用いることでシステムが発話欲求の推定をしやすく、また参加者自身が推定された発話欲求の度合いを調整しやすくなると考えた。参加者に発話欲求があるときに専用のボタンを能動的に押させる方法なども考えられるが、普段の会話で行っていない動作を用いることは、かえって参加者の負担に

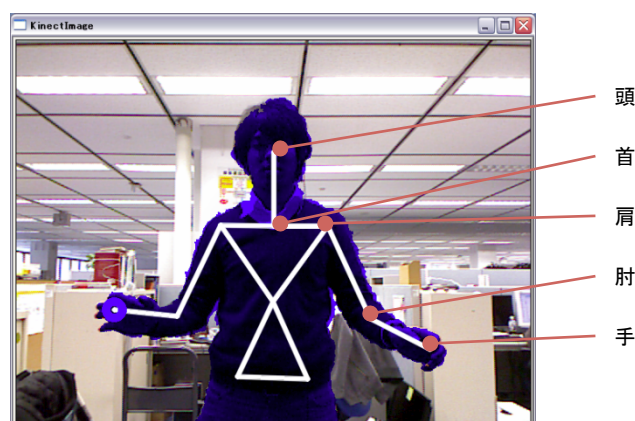


図 6.3: キャプチャされた参加者の関節の3次元位置

なると考えた．Kinect センサ [35] を用い，人を認識し，その頭部，首，肩関節，肘関節，手首などの三次元座標をリアルタイムに取得する（図 6.3）．Kinect センサの画像解像度は 1280×1024 ，奥行き方向の解像度は 640×480 であった．本プロトタイプ実装時点では Kinect センサの画角に参加者の腰より上部が収まらなくては上記三次元座標を安定して取得できなかった．そのため Kinect センサをディスプレイ上部に取り付け，斜め下へ向けることで参加者の腰より上部の RGB 画像，奥行き画像を取得した．それぞれの予備動作候補の検知基準は以下のようにした．

- 手を挙げる

肘関節の位置よりも手首の位置が天地方向で高くなった場合，手を挙げていると判定した（図 6.4(a)）．

- 手を顔周辺へ持っていく

手首と頭部の三次元座標の距離が，ある一定の閾値（今回は 270mm に設定）よりも短くなった場合，手を顔周辺へ持っていく予備動作候補であると判定した

- 頷く

奥行き方向の軸を z 軸とし，Kinect センサから離れる方向を正とする（図 6.4(c)）．頭部の z 座標値を z_h ，首の z 座標値を z_n ，ある時間経過した後の z 座標値をそれぞれ z_h' ， z_n' とすると，

$$z_h - z_n > z_h' - z_n' \quad (6.1)$$

が満たされる場合，頷いていると判定した．

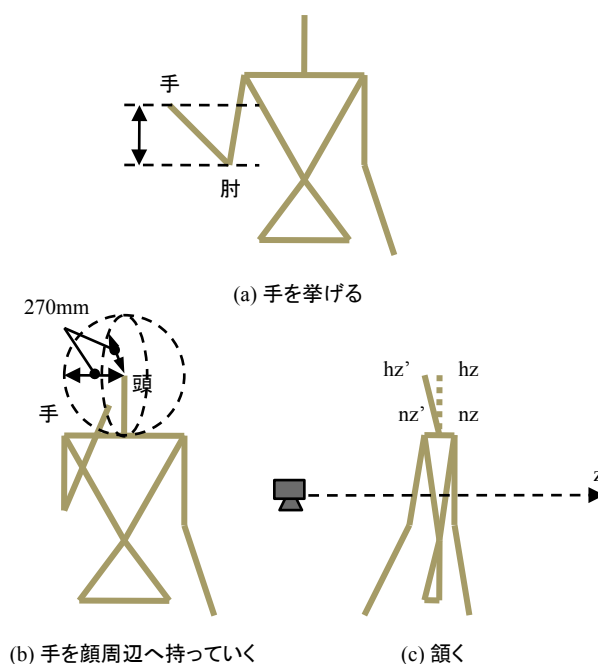


図 6.4: 予備動作の検知方法

6.2.2 発話欲求推定部

前節で検知した予備動作候補を基に，参加者毎に発話欲求フィードバックインジケータを表示した（図 6.5(a)）．予備動作候補ごとに一定の値（スコア）を決め，検知された動作に対応したスコアの総和を発話欲求度とした．発話欲求度は 0 以上の値をとるものとした．発話欲求フィードバックインジケータは，参加者映像下部左側に緑色のゲージとして表示させ，同右側にはマイク入力音声レベルインジケータを赤色で表示させた（図 6.5(c)）．発話欲求度が 0 のとき，ゲージは表示されず，100 のときに丁度左半分を緑色のゲージが埋めるようにし，ゲージの長さで発話欲求度を表現した．発話欲求度が 100 を超えた場合，ゲージは左半分を埋めた状態のままにした．また，マイクから入力音声がある場合，発話欲求度は常に 100 となるようにした．発話欲求が 100 になり満たされてから発話にいたる（音圧が 1 以上の状態）までを連続的に見せることで，参加者が「発話欲求のある状態」を自然に認知できると考えた．発話欲求度は時間経過とともに一定の速度で減少させるようにした．

本手法において参加者にとって意味があるのは，発話欲求度が 50，60 などの具体的な数値ではなく，100 に達しているか否か，100 に達しそうであるか否かである．参加者はこれを見て発話欲求の有無に応じて動きを制御し，発話欲求度を 100 まで上げるか否かを選択する．上記，予備動作候補ごとに設定されたスコアの大きさは，会議実施前に参加者が，予備動作をしてから発話することを数回繰り返して，微調整できるようにした．

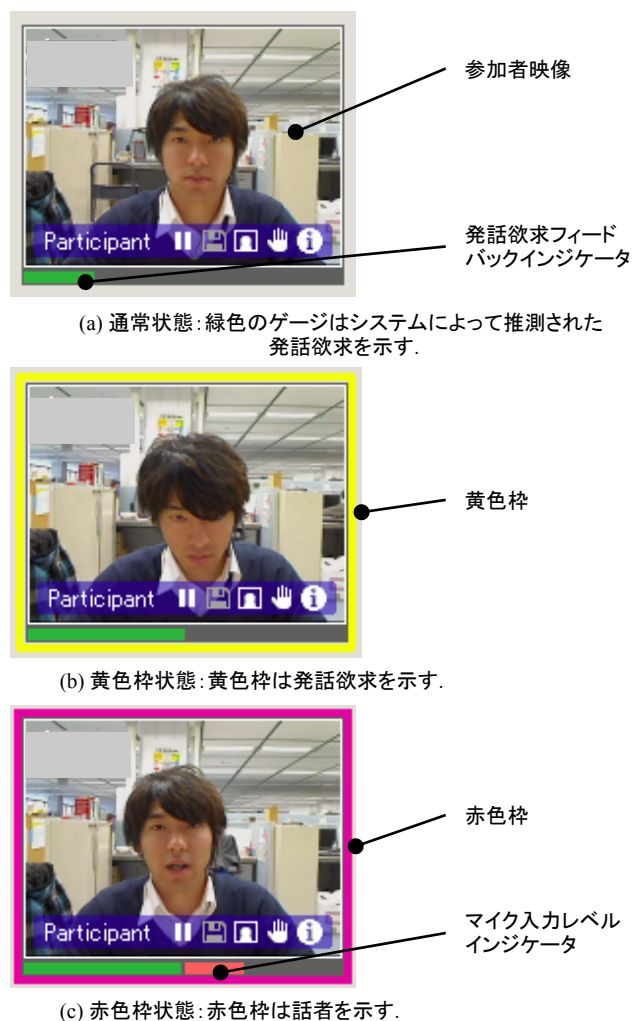


図 6.5: 参加者映像

6.2.3 発話欲求伝達部

ある参加者の発話欲求度が100を超えたときに、その参加者映像の枠を黄色く強調した(図6.5(b))。提案機能を実装している商用システムのMeetingPlaza[39]は、発話中の参加者映像に赤色枠を表示する仕様となっている。発話の前段階の黄色から、発話中を赤色へ変化することは、人が普段慣れ親しんでいる信号機の色の変化と等しく、参加者が状態の遷移をイメージしやすくなると考えた。

6.3 評価実験

6.3.1 概要

この実験の目的は、本プロトタイプにより、発話欲求を正しくシグナルへ反映し、伝達でき、発話衝突を低減できるかどうか、ユーザ評価実験にて確かめることである。具体的には、3人の被験者にWeb会議システムを用いてアイデアを創造する会議を行い、発話欲求を伝達し、発話衝突確率が低減できるかを検証した。

比較条件として、通常のWeb会議システムを使用した条件と、発話欲求を伝えるために「ボタン」を使用した条件とを用意した。ボタン条件は、意識的な操作により発話欲求を伝達する単純な方法であるため、比較対象として選定した。

6.3.2 手順

20代男女9人の被験者を会議の参加者とした。被験者は3人一組のグループに分かれた。グループを構成する被験者は、互いに親しい間柄であった。遠隔会議システムにはMeetingPlaza[39]を使用した。被験者はそれぞれ、防音室に入り、クライアント端末からMeetingPlazaにログインした。会議参加者の映像は、2人分を横に並べて配置し、もう1人を下に配置した。すべての参加者は、マイクとヘッドホンの組み込まれたヘッドセットを装着した。いずれかの参加者のマイクに音声入力があった場合、その参加者映像に赤色の枠が表示されることで、その参加者が発話中であることがすべての参加者へと伝えられた。

下記3つの実験条件を用意した。

- (1) Web会議条件：上記話者を知らせる機能のみを使用し、参加者には特別な操作を要求しない。
- (2) ボタン制御条件：提案手法と比較するため、単純な操作で明確な発話欲求を伝えることができる仕組みのうち、最もシンプルだと考えられる条件を用意した。上記話者を知らせる機能に加え、参加者が発話したいときにキーボードのスペースキーを押下することで、その参加者の映像に黄色枠を表示できる機能を持たせた。
- (3) 提案手法条件：上記話者を知らせる機能に加え、発話欲求伝達モジュールを使用し、発話欲求フィードバックインジケータと、発話欲求を他の参加者へ伝えるための黄色枠表示機能を持たせた。被験者には実験前に、検知される予備動作の種類を教示し、黄色枠を使用して発話欲求を伝えるよう指示した。図1のように、画面右側に、自身の映像と、予備動作が検知されているか否かをモニタリングできるウィンドウを表示した。

条件(2)と(3)では、被験者へ黄色枠の表示されるルールを説明し、「黄色枠を活用して話者交替を行ってください」と指示した。本実験では提案手法の機能を用いて会話するこ

表 6.1: 被験者グループごとの実施した条件順序

グループ	1 回目	2 回目	3 回目
グループ 1	条件 1	条件 2	条件 3
グループ 2	条件 2	条件 3	条件 1
グループ 3	条件 3	条件 1	条件 2

とでどの程度発話衝突を低減できるかという点を評価することが目的であるため、黄色枠を活用するよう指示した。今後実用性評価をする段階においては、被験者へ黄色枠の表示されるルールのみを伝えて実験を行い、提案機能がどの程度利用されるかを調査することも必要であると考え、上記3つの条件にて、8分間ずつ会話をし、操作に慣れてから実験に臨んだ。会議の内容は、携帯電話の新機能のアイデア出しをすることであった。被験者は30分間議論を行い、10分ごとに条件を変更した（表1）。著者はこれまでに類似の実験を行い、1時間未満の会議において、時間経過とともに発話回数が減少しないことを確認している[60][61]。本実験においても3通りの組み合わせで実験を行い、順序効果がないことを確認し、実験を終了した。各条件実施直後に、被験者に「発話が衝突したと感じた度合い」を5段階で回答させた。「発話がよく衝突した」と感じた場合は5、そうでない場合は1とし、5段階のスコアをつける主観評価を行った。計測用にクライアント端末を1台用意し、MeetingPlazaへログインし、実験の様子を、実験者2人が観察した。1人の実験者が会話のターン数をカウントし、もう1人が発話衝突回数をカウントした。会議中の音声を聞きながら、話者が切り替わったときにターンが遷移したと判定し、また、2人以上の参加者が同時に発話を開始したときに発話衝突が生じたと判定した。

6.3.3 結果

ある話者が発話開始してから発話終了するまでをターンと呼ぶこととし、実験条件ごとのターン数を図6.6に示す。Web会議条件で、最もターン数が多く、次が提案手法条件となった。一元配置分散分析を行ったところ、群間自由度2、群内自由度6、P値0.869であり、ターン数に統計的な差異は認められなかった。

条件ごとの発話衝突確率を図6.7に示す。本稿では、発話衝突回数をターン数で割った値を発話衝突確率とした。発話衝突確率はWeb会議条件で最も高く、提案手法条件で最も低くなった。一元配置分散分析を行ったところ、群間自由度2、群内自由度6、P値0.099であり、有意水準1%で、条件により発話衝突確率が異なるという結果が得られた。Bonferroni-Holmの方法により有意水準を補正し、t検定を用いて多重比較したところ、Web会議条件よりも、提案手法条件の発話衝突確率の方が、有意水準5%（補正後の有意水準は1.67%）で有意に低いという結果が得られた。

続いて「発話が衝突したと感じた度合い」を回答させた主観評価の結果を図6.8に示す。Friedman検定を実施したところ、自由度2、カイ二乗値8.07、有意確率（P値）0.018

(< 0.05) となり，会議条件ごとに主観評価結果に差があることが分かった．そこでさらに Bonferroni-Holm の方法による有意水準の補正を行い，Wilcoxon の符号順位検定を用いて多重比較を行ったところ，Web 会議条件は提案手法条件と比較し，有意差 5 % で ($p < .0167$) 「発話がよく衝突した」と被験者が感じたことが分かった．

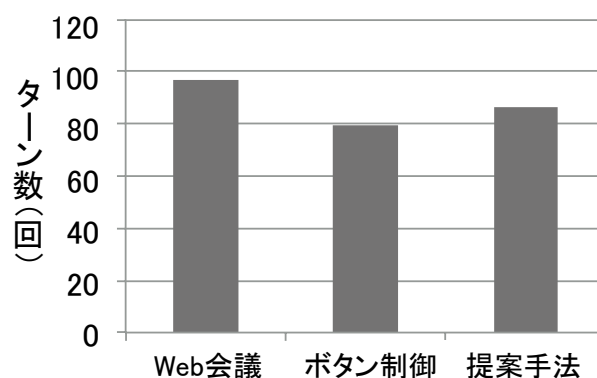


図 6.6: ターン数の比較

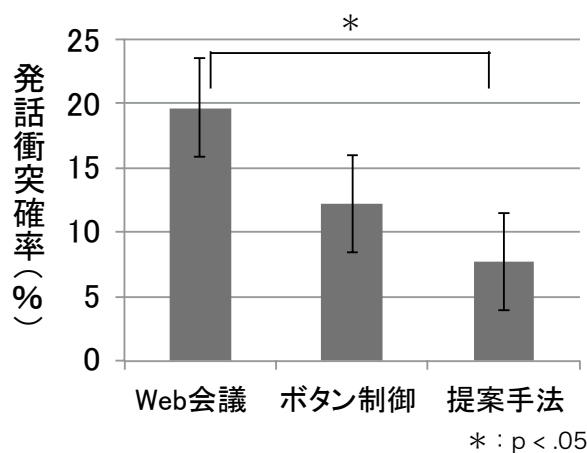


図 6.7: 発話衝突確率の比較

6.3.4 考察

Web 会議条件，ボタン制御条件，提案手法条件のうち，提案手法条件での発話衝突確率が最も低かった．Web 会議条件と比較すると平均値が半分以下であり，統計的に有意

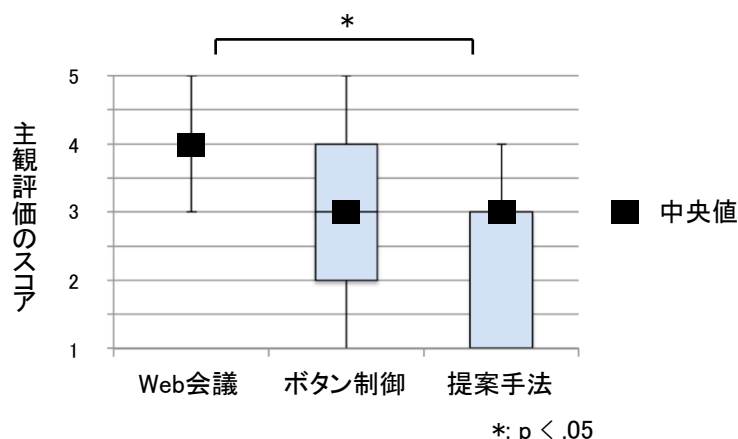


図 6.8: 「発話が衝突したと感じた度合い」の比較

な差がみられた。このことから、提案手法により、発話欲求を伝達して発話衝突を低減することができたといえる。また主観評価の結果から、被験者が提案手法条件では Web 会議条件よりも発話衝突が少ないと感じたことが分かった。本実験結果ではボタン制御条件と提案手法条件の間に統計的差異は認められなかったが、定量的評価と主観評価両方において、提案手法条件で最も良好な結果が得られた。

ボタン制御条件は最も明示的かつ単純に発話欲求を伝達できる条件であったが、Web 会議条件と比較して、定量評価、主観評価結果ともに有意な差がみられなかった。録画したビデオを解析したところ、ボタン制御条件ではボタンを押して発話したにも関わらず発話衝突する場面が多く見られた。ボタン制御条件ではボタンを押せばすぐに黄色枠が表示されるため、ボタンを押した後にすぐ発話することが多くみられた。一方「挙手」は自分のインジケータがたまるのを待つ間に、他の参加者に黄色枠が表示されるか否かを見て、確認してから発話する様子がみられた。ある参加者に発話欲求が生じた際に、他の参加者に黄色枠が表示されるか否かを確認する程度が異なることが、提案手法条件とボタン制御条件における発話衝突を低減する効果の違いに現れた可能性が考えられる。

ビデオ解析によると、被験者は提案手法条件においてほとんどの発話の前に何らかの動作をし、発話欲求度を 100 にし、黄色枠を表示させてから発話する、という手順を踏んでいた。このことから被験者は自身に発話欲求があるときに、その発話欲求とシステムに検知されている発話欲求度との整合をとっていたことが分かる。また、黄色枠が表示されたことに気づくのが遅くて発話衝突する場面は観察されたが、黄色枠を無視して発話することはなかった。

提案手法条件やボタン制御条件では被験者に特別な操作を要求したため負荷がかかり、結果的に著しくターン数が減少するなどの可能性も考えられたが、実験結果に差は見られなかった(図 6.6)。しかしある被験者から「発話前に特別な動作を挟むことが自然な発話開始の弊害となる」というコメントが得られた。この点は次に解決すべき課題である。

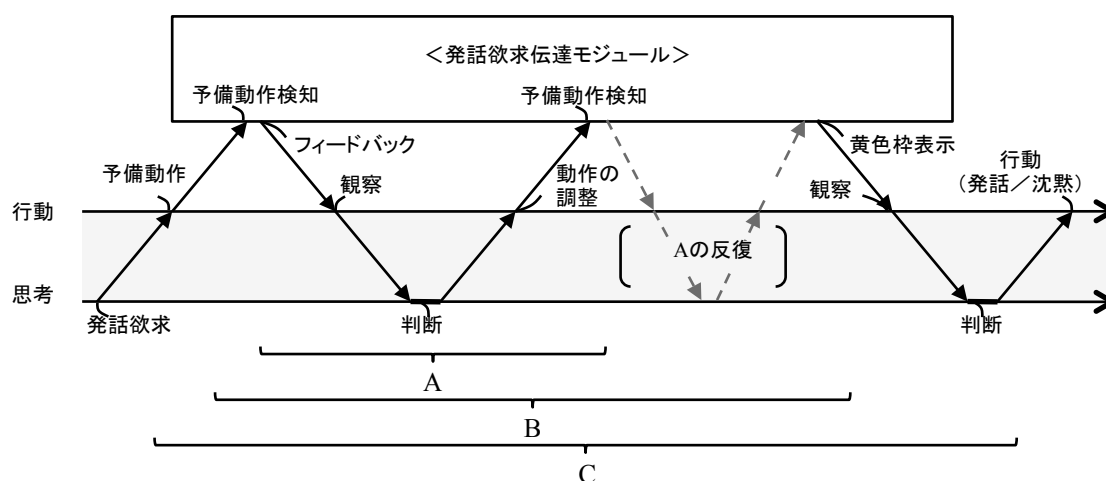


図 6.9: 話者交替において提案手法を適用した際の認知・行動モデル

話者交替において、本提案手法を用いた場合の認知・行動モデルは図 4.6 にいくつかのステップを追加し、図 6.9 のようなモデルになるという仮説を立てた。具体的には、B 間のステップが新たに加わった。参加者はフィードバックインジケータを観察し、自身の動作を調整する。A 間のステップは、フィードバックインジケータと参加者の実際の発話欲求が合致するまで繰り返される。現状のプロトタイプでは、発話欲求が生じてから発話するまでに 1) 対面したコミュニケーション時よりも多くのステップを要すること、2) その操作を行う頻度が高いこと、が参加者へ負担をかけることが考えられる。これらの 2 点の要因による影響を低減するために以下の項目を検討すべきであると考えた。

- (1) 最初の予備動作検知ステップと黄色枠表示ステップの間を短くするために、図 6.9 の A で示されたフィードバックステップと予備動作検知ステップ間の繰り返し回数を減らす。
 - (a) 予備動作に割り当てるスコアを高く設定し、発話欲求フィードバックインジケータのゲージが溜まる速度を速く
 - (b) 予備動作候補を検知した後に発話する回数が多い程、その動作のスコアを高めていく。
- (2) 予備動作ステップと判断ステップ（図 6.9 の C）を通る機会を減らす。すべての話者交替時に、予備動作をして発話欲求を伝達する手順を必要とせず、発話衝突のリスクがある場合のみ、参加者にその予備動作をして発話欲求を表すことを促す。

今後はこれらの方法を具体的に検討していき、Web 会議において参加者へ負荷をかけずに発話衝突をさらに低減し、自席のデスクトップパソコンやノートパソコンからでも、創造会議を快適に行える遠隔会議環境の実現を目指す。

第7章 結論

本研究は、Web 会議での発話衝突を低減することにより、自席のデスクトップパソコンやノートパソコンからでも、創造会議を快適に行える遠隔会議環境を実現することを目的とした。

発話衝突を低減させる手法を模索するために、まず発話衝突の原因とその影響について調査した。話者交替時の認知・行動モデルを定義し、それがどのように音声遅延の影響を受けて発話衝突が生じるかという仮説をたて、実験を行った。音声会議環境において音声遅延量を変化させ、発話衝突確率と、参加者の受ける精神的ストレスを測定した結果、音声遅延量が 400msec を超える場合には発話衝突確率が増大し、精神的ストレスが高まる結果となった。したがって、本研究が対象とする Web 会議の環境では音声遅延量が 400msec を超えることが多いため、発話衝突確率を低減させることが望まれた。

そこで本論文では、次話者候補提示手法と発話欲求伝達手法から構成される、遠隔コミュニケーションにおける発話衝突低減手法を提案した。

次話者候補提示手法は、システムが会議参加者各々の予備動作を検知し、その種類と頻度から、次に最も発話しそうな参加者を次話者候補として選定し、すべての参加者へ伝達する手法である。システムは最も多く予備動作を行った参加者を、そのときの次話者候補として選定する。そして、誰が次話者候補であるのかをすべての参加者へ知らせるため、選定された参加者の映像を強調して表示させる。この手法を実現するプロトタイプシステムを実装して評価実験を行った結果、本提案手法により Web 会議において発話衝突確率を低減できる可能性が示された。

発話欲求伝達手法は、システムによって推定された発話欲求の度合いを随時参加者に提示することにより、参加者自身に予備動作を調整させ、発話欲求の推定精度を高める手法である。システムが参加者の発話欲求を正確に推定することはできないため、参加者がそれを補助することで推定精度を高められるようにした。この手法を実現するプロトタイプシステムを実装して評価実験を行い、本提案手法を用いることで通常の Web 会議と比較して発話衝突確率を半分に以下に低減させられることを確認した。

以上の結果から、本論文で提案した発話衝突低減手法により、Web 会議において発話衝突確率が低減できることが分かった。したがって、複数の離れた拠点にいる参加者が自席のデスクトップパソコンやノートパソコンから参加し、創造会議を快適に行える遠隔会議環境を実現できる可能性を示すことができた。

本研究の学術的な貢献は、遠隔コミュニケーションにおける話者交替時の認知・行動モデルを定義したことと、遠隔コミュニケーションにおいて話者交替時の発話衝突を低減するための手法を提案したことである。これまで対面コミュニケーションにおいては話者交替のルールが定義され、それに基づいて多くの議論がなされていた。しかし遠隔コミュニケーションにおいてそれらのルールがどのように変化するか整理し、それを基に遠隔コミュニケーションならではの話者交替の問題を議論した例は少なかった。本研究において遠隔コミュニケーションにおける話者交替時の認知・行動モデルを定義したことで、今後発話衝突に限らず、話者交替の他の問題に関してもこれに基づいて議論や研究を進めていくことができる。発話衝突から復帰する、発話権の譲渡を円滑にする、意図せぬ沈黙を防ぐなどの課題の解決方法を考える際にも、このモデルのどのステップに問題が生じている

かを捉え、どのようにそれを解決するかを筋道立てて考えることができる。

さらに、自席のデスクトップパソコンやノートパソコンからでも実施可能な Web 会議システムで創造会議を行う際の大きな弊害となっている発話衝突を解消する手法を提案したことで、導入や利用のために必要な機材が少なく、利用する場所の自由度が高く、さらに会話のしやすい環境を実現するための大きな一歩を示すことに貢献した。

謝辭

本研究は非常に多くの方々のご支援のもと行われました。

学部時代から合わせて7年もの間、終始格別なご指導を賜りました、慶應義塾大学情報工学科 岡田謙一教授に謹んで深い感謝の意を表します。また、研究内容に関するご議論と、繰り返し論文執筆方法のご指導を賜りました、慶應義塾大学情報工学科 萩原将文教授、重野寛教授、斎藤博昭准教授に心より感謝申し上げます。

本研究の企画から進め方までご指導ご鞭撻をいただきました、NTT サービスエボリューション研究所 東野豪氏、小林稔氏、井原雅行氏、中茂睦裕氏に心より感謝申し上げます。在職中に学位を取得することにご理解、ご協力いただきました NTT コミュニケーションズ株式会社 貞田洋明氏、井藤雅稔氏にも感謝申し上げます。

そして、大学学部時代から現在まで、研究活動ならびに学位取得に際して特にお力添えいただいた坂内祐一氏、川口信隆氏、上田真太郎氏、宮田章裕氏、宮挟和大氏、鈴木雄士氏、山本俊氏、岡島雄太氏、吉川誠氏に深く感謝いたします。博士課程をともに闘い、多大なるご支援をいただいた鈴木正基氏、上条亜美氏、遠藤伶氏にも心から感謝いたします。

最後に、これまでの研究生生活を支えてくれた家族に感謝の意を表します。とりわけ、挙式後、落ち着いた時間を全く過ごせなかったにもかかわらず、本論文執筆を全面的に支えてくれた妻 千尋に心から感謝します。感謝すべき全ての方をここに記すことはできませんが、私にとって家族、友人、先輩、後輩、関わった全ての方が指導者です。ここに謹んで感謝の意を表します。

2013 年 2 月
玉木 秀和

参考文献

- [1] Birdwhistell, R. L. *Kinesics and Context: Essays on Body Motion Communication*. Univ of Pennsylvania Press, 1970.
- [2] Burke, J. and Murphy, R. Rsvp: An investigation of remote shared visual presence as common ground for human-robot teams. *HRI '07*, pp. 161–168, 2007.
- [3] Buxton, B. Telepresence: Integrating shared task and personal spaces. In *In Proceedings of The Potential of Team and Organizational Computing(Groupware '91)*, pp. 27–36, 10 1991.
- [4] Cassell, J. and Vilhjálmsón, H. Fully embodied conversational avatars: Making communicative behaviors autonomous. *Autonomous Agents and Multi-Agent Systems*, Vol. 2, pp. 45–64, 1999.
- [5] Chidambaram, V., Chiang, Y.-H., and Mutlu, B. Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In *HRI '12: Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pp. 293–300, 2012.
- [6] Telepresence シスコシステムズ合同会社. <http://www.cisco.com/en/US/products/ps7060/index.html>, 2012.
- [7] Cooley, C. H. *Social Organization: A Study of Larger Mind*. Charles Scribner's Sons, 1909.
- [8] Czerwinski, M., Lund, A., and Tan, D. A comparison of chat and audio in media rich environments. In *CHI EA '08: CHI '08 Extended Abstracts on Human Factors in Computing Systems*, pp. 323–332, 2008.
- [9] Dance, F. E. X. and Larson, C. E. *The functions of human communication : a theoretical approach*. New York: Holt, Rinehart and Winston, 1976.
- [10] Dong, W. and Fu, W.-T. One piece at a time: Why video-based communication is better for negotiation and conflict resolution. *CSCW '12*, pp. 167–176, 2012.

- [11] 榎本美香. 日本語における聞き手の話者移行適格場の認知メカニズム. ひつじ書房, 2009.
- [12] 深田博己. インターパーソナル・コミュニケーション-対人コミュニケーションの心理学-. 北大路書房, 1995.
- [13] Hall, E. T. *The Hidden Dimension*. Doubleday and Co., 1966.
- [14] 原岡一馬. 人間とコミュニケーション. ナカニシヤ出版, 1990.
- [15] Hawkey, K., Kellar, M., Reilly, D., Whalen, T., and Inkpen, K. M. The proximity factor: impact of distance on co-located collaboration. In *GROUP '05: Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, pp. 31–40, 2005.
- [16] Heeler, P. and Hardy, C. A preliminary report on the use of video technology in online courses. *Journal of Computing Sciences in Colleges*, Vol. 20, No. 4, pp. 127–133, 2005.
- [17] Hiltz, S. R. and Turoff, M. Structuring computer-mediated communication systems to avoid information overload. *Communications of the ACM*, Vol. 28, No. 7, pp. 680–689, 1985.
- [18] 井上智雄, 岡田謙一, 松下温. テレビ会議における映像表現の利用とその影響. 情報処理学会論文誌, Vol. 40, No. 10, pp. 3752–3761, 1999.
- [19] インタロボット株式会社. *InterRobot*. <http://www.i-robot.co.jp>, 2013.
- [20] Isaacs, E. A., Morris, T., and Rodriguez, T. K. A forum for supporting interactive presentations to distributed audiences. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pp. 405–416, 1994.
- [21] 石井亮, 宮島俊光, 藤田欣也. アバタ音声チャットシステムにおける会話促進のための注視制御. ヒューマンインタフェース学会論文誌, Vol. 10, No. 1, pp. 87–94, 2008.
- [22] 石井裕, 渡辺富夫. Virtualactor を対面合成した身体的ビデオコミュニケーションシステム. ヒューマンインタフェース学会論文誌, Vol. 5, No. 2, pp. 225–234, 5 2003.
- [23] ITU-T. *One-way transmission time*. ITU-T Rec. G.114, 2003.
- [24] Kawashima, H., Nishikawa, T., and Matsuyama, T. Visual filler: Facilitating smooth turn-taking in video conferencing with transmission delay. In *CHI 2008 Extended Abstract*, pp. 3585–3590, 2008.

- [25] Kelley, J. F. CAL - a natural language program developed with the OZ paradigm: Implications for supercomputing systems. In *First International Conference on Supercomputing Systems*, pp. 238–248, 1985.
- [26] Kim, T. J., Chang, A., Holland, L., and Pentlnd, A. S. Meeting mediator: Enhancing group collaboration with sociometric feedback. In *CHI 2008 Proceedings Works In Progress*, pp. 3183–3188, 2008.
- [27] Koh, E. Conferencing room for telepresence with remote participants. In *GROUP '10: Proceedings of the 16th ACM international conference on Supporting group work*, pp. 309–310, 2011.
- [28] 小磯花絵, 伝康晴. 円滑な話者交替はいかにして成立するか：—会話コーパスの分析にもとづく考察—. 認知科学, Vol. 7, pp. 93–106, 2000.
- [29] Lopez, B., Hernandez, A., Fernandez, R., and Hernandez, L. Design and validation of eca gestures to improve dialogue system robustness. In *Proceedings of the Workshop on Embodied Language Processing*, pp. 67–74, 2007.
- [30] Lucia, A. D., Francese, R., Passero, I., and Tortora, G. SLMeeting: supporting collaborative work in Second Life. In *AVI '08: Proceedings of the working conference on Advanced visual interfaces*, pp. 301–304, 2008.
- [31] 松田晃一, 三宅貴浩. パーソナルエージェント指向仮想社会 paw (第2版) の構築と評価. 情報処理学会論文誌, Vol. 41, No. 10, 2000.
- [32] 松尾隆. コミュニケーションの心理学. ナカニシヤ出版, 1999.
- [33] 松下温, 岡田謙一. コラボレーションとコミュニケーション. 共立出版, 1995.
- [34] Mehrabian, A. *Silent messages*. Wadsworth Publishing Company, 1971.
- [35] Microsoft. Xbox Kinect センサ. <http://www.xbox.com/ja-JP/kinect>, 2012.
- [36] Mutlu, B., Kanda, T., Forlizzi, J., Hodgins, J., and Ishiguro, H. Conversational gaze mechanisms for humanlike robots. Vol. 1, No. 2, pp. 1–33, 2012.
- [37] 中茂睦裕, 中山彰, 小林稔. テーブルトップへの画像刺激による遠隔コミュニケーションの分析. ヒューマンインタフェースシンポジウム 2006 論文集 2, pp. 695–700, 2006.
- [38] 新徳健, 渡辺富夫. 3 者間インタラクション支援および合成的解析のための身体的バーチャルコミュニケーションシステム. 情報処理学会論文誌, Vol. 45, No. 4, pp. 1212–1221, 4 2004.

- [39] NTT アイティ株式会社. *MeetingPlaza*. <http://www.meetingplaza.com/index-j.html>, 2012.
- [40] Nunamaker, J. F., Dennis, A. R., Valacich, J. S., Vogel, D., and George, J. F. Electronic meeting systems. *Communications of the ACM*, Vol. 34, No. 7, pp. 40–61, 1991.
- [41] Ocker, R. J. and Fjermestad, J. Communication differences in virtual design teams: Findings from a multi-method analysis of high and low performing experimental teams. *The DATA BASE for Advances in Information Systems*, Vol. 39, No. 1, pp. 51–67, 2008.
- [42] O'hara, K., Kjeldskov, J., and Paay, J. Blended interaction spaces for distributed team collaboration. *Transactions on Computer-Human Interaction (TOCHI)*, Vol. 18, No. 1, p. Article 3, 2011.
- [43] Okada, K., Maeda, F., Ichikawaa, Y., and Matsushita, Y. Multiparty videoconferencing at virtual social distance: Majic design. In *CSCW '94: Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pp. 385–393, 1994.
- [44] 岡田謙一, 西田正吾, 葛岡英明, 仲谷美江, 塩澤秀和. ヒューマンコンピュータインタラクション. オーム社, 2002.
- [45] 大崎浩司, 渡辺富夫, 山本倫也. Interactor に手指動作入力を併用した身体的インタラクションシステムの開発. ヒューマンインタフェース学会論文誌, Vol. 7, No. 3, pp. 89–98, 2005.
- [46] Park, Y., Bae, S., and Nam, T. How do couples use cheektouch over phone calls? In *CHI 2012*, pp. 763–766, 2012.
- [47] Patterson, M. L. *Nonverbal Behavior, A Functional Perspective*. Springer-Verlag, 1983.
- [48] Regenbrecht, H., Ott, C., Wagner, M., Lum, T., Kohler, P., Wilke, W., and Mueller, E. An augmented virtuality approach to 3d videoconferencing. In *Proceedings of the The 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR2003)*, pp. 290–291, 2003.
- [49] Richmond, V. P. and McCroskey, J. C. *Nonverbal Behavior in Interpersonal Relations*. Allyn and Bacon, 2003.
- [50] Sacks, H., Schegloff, E. A., and Jefferson, G. A simplest systematics for the organization of turn-taking for conversation. *Language*, Vol. 50, No. 4, pp. 696–735, 1974.

- [51] Sanchez-Cortes, D., Jayagopi, D. B., and Gatica-Perez, D. Predicting remote versus collocated group interactions using nonverbal cues. ICMI-MLMI '09 Workshop on Multimodal Sensor-Based Systems and Mobile Phones for Social Computing, 2009.
- [52] Schramm, W. How communication works. In *The Processes and Effects of Mass Communication*, pp. 3–26, 1954.
- [53] Scissors, L., Shami, N. S., Ishihara, T., Rohall, S., and Saito, S. Real-time collaborative editing behavior in. In *CHI 2011*, pp. 1119–1128, 2011.
- [54] シードプランニング. 2008 年版テレビ会議 / Web 会議の最新市場と HD 化動向. シードプランニング, 2008.
- [55] Sellen, A. J. Speech patterns in video-mediated conversations. CHI '92, pp. 49–59, 1992.
- [56] 立川敬二. コミュニケーションの構造. NTT 出版株式会社, 1993.
- [57] Tada, H. Eyeblick rates as a function of the interest value of video stimuli. In *Tohoku Psychologica Folia*, Vol. 45, pp. 107–113, 1986.
- [58] 田多英興, 山田富美雄, 福田恭介. まばたきの心理学. 北大路書房, 1991.
- [59] 高橋誠. 会議の進め方. 日本経済新聞出版社, 1987.
- [60] 玉木秀和, 中茂睦裕, 東野豪, 小林稔. 人のコミュニケーションリズムに着目した web 会議円滑化手法. In *IEICE Technical Report MVE2009*, pp. 101–106, 2009.
- [61] 玉木秀和, 中茂睦裕, 東野豪, 小林稔. 遠隔会議における話者交替円滑化手法. マルチメディア, 分散, 協調とモバイルシンポジウム DICOMO2010, pp. 108–116, 2010.
- [62] 玉木秀和, 東野豪, 小林稔, 井原雅行. 遠隔会議における発話衝突低減手法. 情報処理学会論文誌, Vol. 53, No. 7, pp. 1797–1806, 2012.
- [63] Tang, J. C. and Rua, M. Montage: Providing teleproximity for distributed groups. In *Proceedings of the Conference on Computer Human Interaction (CHI)'94*, pp. 37–43, 4 1994.
- [64] Tang, J. C. Approaching and leave-taking: Negotiating contact in computer-mediated communication. *Transactions on Computer-Human Interaction*, Vol. 14, No. 1, pp. 1–26, 2007.
- [65] 徳勲, 友保康成, 渋谷雄, 田村博. テレビ会議技術の課題と利用法についての考察. In *8th Symposium on Human Interface*, pp. 207–212, 1992.

- [66] 津田兼六, 鈴木直人. 主観的興味が瞬目率と体動の生起頻度に及ぼす影響. 生理心理学と精神生理学, 第8巻, pp. 31–37, 1990.
- [67] Vargas, M. F. *LOUDER THAN WORDS -An Introduction to Nonverbal Communication-*. Iowa State University Press, 1987.
- [68] 渡辺富夫, 大久保雅史. 身体的コミュニケーション解析のためのバーチャルコミュニケーションシステム. 情報処理学会論文誌, Vol. 40, No. 2, pp. 670–676, 2 1999.
- [69] Watanabe, T., Okubo, M., Nakashige, M., and Danbara, R. Interactor: Speech-driven embodied interactive actor. In *International Journal of Human-Computer Interaction*, pp. 43–60, 3 2004.
- [70] Whittaker, S., Swanson, J., Kucan, J., and Sidner, C. Telenotes: managing lightweight interactions in the desktop. *June 1997 Transactions on Computer-Human Interaction (TOCHI)*, Vol. 4, No. 2, pp. 137–168, 1997.
- [71] Wood, J. T. *Human communication : A symbolic interactionist perspective*. Holt, Rinehart, and Winston, 1982.
- [72] Yamashita, N., Kuzuoka, H., Hirata, K., Aoyagi, S., and Shirai, Y. Supporting fluid tabletop collaboration across distances. In *CHI 2011 Session: Tabletop Synchronous Collaboration*, pp. 2827–2836, 2011.
- [73] 鎧沢勇, 滝川啓, 大久保榮, 渡辺義郎. 衛星通信を利用した画像会議におけるエコー及び伝搬遅延の影響. 電子通信学会論文誌 (B), Vol. 64B, No. 11, pp. 1281–1288, 1981.
- [74] 由井園隆也, 宗森純. 発想支援ブルーウェア郡元の効果～数百の試用実験より得たもの～. 人工知能学会論文誌, Vol. 19, No. 2, pp. 105–112, 2004.
- [75] Zinn, K. L., Parnes, R., and Hench, H. Computer-based educational communications at the university of michigan. In *Proceedings of the annual conference*, pp. 150–154, 1976.

論文目録

主論文に関する原著論文

- (1) 玉木秀和，東野豪，小林稔，井原雅行，岡田謙一．：遠隔会議における発話衝突低減手法．情報処理学会論文誌 53(7), pp.1797–1806, 2012 年 7 月．
 - (2) 玉木秀和，東野豪，小林稔，井原雅行．：音声遅延が遠隔会議中の発話衝突と精神的ストレスに与える影響．電子情報通信学会論文誌 Vol.J96-D, No.1, 2013 年 1 月（掲載予定）．
 - (3) 玉木秀和，東野豪，小林稔，井原雅行．：発話がぶつからない Web 会議を実現するための発話欲求伝達手法，情報処理学会論文誌, 2013 年 1 月（掲載予定）．
-

国際会議

- (1) Hidekazu Tamaki, Suguru Higashino, Minoru Kobayashi, Masayuki Ihara, :Reducing Speech Contention in Web Conferences, IEEE SAINT 2011, pp.75–81, 2011.
-

研究会

- (1) 玉木秀和，中茂睦裕，東野豪，小林稔，鈴木由里子，：人のコミュニケーションリズムに着目した映像通信メディア，情報処理学会グループウェアとネットワークサービスワークショップ 2009, 2009 年 9 月．
- (2) 玉木秀和，中茂睦裕，東野豪，小林稔，：人のコミュニケーションリズムに着目した Web 会議円滑化手法，情報処理学会研究報告 MVE 2009, pp.101–106, 2009 年．
- (3) 玉木秀和，東野豪，小林稔，井原雅行，：遠隔会議における話者交替円滑化手法，マルチメディア，分散，協調とモバイルシンポジウム DICOMO 2010, pp.108–116, 2010 年 7 月．

- (4) 玉木秀和，東野豪，小林稔，井原雅行，：話者交替時に使用される非言語情報は会議環境により変化するか，情報処理学会グループウェアとネットワークサービスワークショップ 2010, 2010 年 9 月．
 - (5) 玉木秀和，東野豪，小林稔，井原雅行，：Web 会議における話者交替円滑化手法の検討，画像電子学会 VMA 研究会, pp.9–18, 2011 年 1 月．
 - (6) 玉木秀和，東野豪，小林稔，井原雅行，：遠隔会議における発話の衝突と精神的ストレスの関係，情報処理学会研究報告 2011-GN-79(10), pp.1–6, 2011 年 3 月．
 - (7) 玉木秀和，東野豪，小林稔，井原雅行，：Web 会議における発話衝突回避のための発話欲求伝達手法，情報処理学会グループウェアとネットワークサービスワークショップ 2011, 2011 年 11 月．
-

受賞

- (1) 情報処理学会 グループウェアとネットワークサービス研究会ワークショップ 2010 ベストプレゼンテーション賞．
 - (2) 情報処理学会 グループウェアとネットワークサービス研究会ワークショップ 2011 ベストプレゼンテーション賞．
 - (3) 情報処理学会 平成 24 年度 山下記念研究賞．
-

その他の活動

論文誌

- (1) 宮狭和大，坂内祐一，鈴木雄士，玉木秀和，重野寛，岡田謙一，：MR 空間における仮想シールを介したシンタックスの異なる実物体の遠隔共有手法，情報処理学会論文誌 48(1), pp.134–147, 2007 年 1 月．
- (2) 坂内祐一，玉木秀和，鈴木雄士，重野寛，岡田謙一，：実物体を用いた MR 空間での遠隔協調作業，情報処理学会論文誌 48(7), pp.2465–2476, 2007 年 7 月．
- (3) 玉木秀和，山本峻，岡嶋雄太，坂内祐一，岡田謙一，：MR 空間における準同期的な実物体共有による遠隔作業支援，日本バーチャルリアリティ学会論文誌 12(4), pp.529–536, 2007 年 12 月．

国際会議

- (1) Hidekazu Tamaki, Suguru Higashino, Minoru Kobayashi, Masayuki Ihara, :Do They Use Different Set of Non-verbal Language in Turn-Taking in Distributed Conferences?, HCI (22), pp.539–543, 2011.
- (2) Hidekazu Tamaki, Yuichi Bannai, Yuji Suzuki, Hiroshi Shigeno, Ken-ichi Okada, :“Remote Collaboration Based on Portable Object in Mixed Reality”, In Proc. of The Second International Conference on Collaboration Technologies, pp.84–89, July, 2006.
- (3) Yuichi Bannai, Hidekazu Tamaki, Yuji Suzuki, Hiroshi Shigeno, Ken-ichi Okada, :A Tangible User Interface for Remote Collaboration System Using Mixed Reality, ICAT 2006, pp.143–154.
- (4) Hidekazu Tamaki, Yuichi Bannai, Shun Yamamoto, Yuta Okajima and Ken-ichi Okada, :“A Sharing Method of Real Objects Differ in Syntax each other through Virtual Sticker between Distant Mixed Reality Spaces”, International Workshop on Informatics 2007, pp.11–17, September, 2007.
- (5) Shun Yamamoto, Hidekazu Tamaki, Yuta Okajima, Yuichi Bannai, Kenichi Okada, :“Symmetric Model of Remote Collaborative MR Using Tangible Replicas”, IEEE-VR 2008, pp.71–74, 2008.
- (6) Shun Yamamoto, Hidekazu Tamaki, Yuichi Bannai, Kenichi Okada, :“Method for Sharing Real Objects with Different Syntax through Virtual Stickers between Distant Mixed Reality Spaces”, International Journal of Informatics Society, Vol.1, No.1, pp.27–34, Jan. 2009.
- (7) Mutsuhiro Nakashige, Minoru Kobayashi, Yuriko Suzuki, Hidekazu Tamaki, Suguru Higashino, :“Hiya-Atsu” media: augmenting digital media with temperature, CHI Extended Abstracts 2009, pp.3181–3186.

学会誌

- (1) 玉木秀和, 東野豪, : インタラクティブメディア・インタフェース, 画像電子学会誌 39(6), pp.813–814, 2010 年 11 月 .

研究会

- (1) 玉木秀和，坂内祐一，宮狭和大，鈴木雄士，重野寛，岡田謙一，：MR 空間における実物体を介した遠隔協調作業，電子情報通信学会技術研究報告 MVE 105(256), pp.29–34, 2005 年 8 月．
 - (2) 山本峻，坂内祐一，鈴木雄士，玉木秀和，岡田謙一，：遠隔協調 MR における実物体インタラクション機能の実現，情報処理学会研究報告 GN 2007(6), pp.19–24, 2007 年 1 月．
 - (3) 玉木秀和，坂内祐一，山本俊，岡嶋雄太，岡田謙一，：MR 空間における準同期的な実物体共有遠隔作業支援，情報処理学会研究報告 GN 2007(56), pp.1–6, 2007 年 6 月．
 - (4) 山本峻，坂内祐一，玉木秀和，岡嶋雄太，岡田謙一，：MR 空間における遠隔間での実物体を用いた実物体の多重化，電子情報通信学会技術研究報告 MVE 107(242), pp.85–90, 2007 年 9 月．
 - (5) 玉木秀和，中茂睦裕，鈴木由里子，小林稔，：情報に対する集中度をコントロール可能なスクリーン，情報処理学会グループウェアとネットワークサービスワークショップ 2008，2008 年 11 月．
 - (6) 木村篤信，鈴木由里子，中茂睦裕，玉木秀和，小林稔，：自省的思考を促進するアイデア外化支援システム，電子情報通信学会技術研究報告 MVE 108(328), pp.15–20, 2008 年 11 月．
 - (7) 中茂睦裕，玉木秀和，鈴木由里子，小林稔，東野豪，：みんな集まれバッジによる対話機会創出の検討，情報処理学会研究報告 GN 2009(3), pp.151–156, 2009 年 1 月．
 - (8) 中茂睦裕，玉木秀和，東野豪，小林稔，：非同期コミュニケーションを促進する微笑み通信メディアの評価，情報処理学会研究報告 2010-HCI-139(6), pp.1–5, 2010 年 7 月．
-